

LEVI LOPES TEIXEIRA

PROJEÇÃO DE SÉRIES TEMPORAIS POR MEIO DE UM MÉTODO HÍBRIDO
WAVELET-NEURAL INTEGRADO COM BOOTSTRAP

Tese apresentada ao Programa de Pós-Graduação em Métodos Numéricos em Engenharia – área de concentração: Programação Matemática, setores de Tecnologia e Ciências Exatas da Universidade Federal do Paraná, como requisito parcial à obtenção do grau de Doutor.

Orientador: Prof. Dr. Paulo Henrique Siqueira.

Coorientador: Prof. Dr. Luiz Albino Teixeira Júnior.

CURITIBA
2015

T266p

Teixeira, Levi Lopes

Projeção de séries temporais por meio de um método híbrido wavelet-neural integrado com bootstrap/ Levi Lopes Teixeira. – Curitiba, 2015.
107 f. : il. color. ; 30 cm.

Tese - Universidade Federal do Paraná, Setor de Tecnologia, Programa de Pós-graduação em Métodos Numéricos em Engenharia, 2015.

Orientador: Paulo Henrique Siqueira – Co-orientador: Luiz Albino Teixeira Júnior.

Bibliografia: p. 98-107.

1. Análise de séries temporais. 2. Análise numérica. 3. Redes neurais (Computação). 4. Programação (Matemática). 5. Bootstrap (Estatística). I. Universidade Federal do Paraná. II. Siqueira, Paulo Henrique. III. Teixeira Júnior, Luiz Albino. IV. Título.

CDD: 519.55

TERMO DE APROVAÇÃO

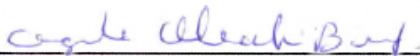
LEVI LOPES TEIXEIRA

PROJEÇÃO DE SÉRIES TEMPORAIS POR MEIO DE UM MÉTODO HÍBRIDO WAVELET-NEURAL INTEGRADO COM *BOOTSTRAP*

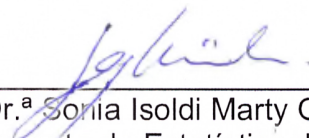
Tese aprovada como requisito parcial para obtenção do grau de Doutor no Curso de Pós-Graduação em Métodos Numéricos em Engenharia – Área de Concentração em Programação Matemática, Setores de Tecnologia e Ciências Exatas da Universidade Federal do Paraná pela banca examinadora:



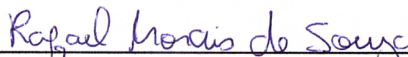
Prof. Dr. Paulo Henrique Siqueira.
Orientador – Departamento Expressão Gráfica, UFPR.



Prof.^a Dr.^a Angela Olandoski Barboza.
Departamento de Matemática, UTFPR.



Prof.^a Dr.^a Sonia Isoldi Marty Gama Müller.
Departamento de Estatística, UFPR.



Prof. Dr. Rafael Moraes de Souza.
Departamento de Ciências Contábeis, UFMG.



Prof.^a Dr.^a Luzia Vidal de Souza.
Departamento Expressão Gráfica, UFPR.

Curitiba, 27 de Agosto de 2015

Dedico este trabalho a Deus,
Senhor de todas as coisas.

Aos meus filhos Vitória,
João Levi e Raquel,
Jóias únicas do meu tesouro.

À minha esposa Lúcia,
Companheira de todas as horas.

AGRADECIMENTOS

Aos professores Paulo Henrique Siqueira e Luiz Albino Teixeira Junior pelas orientações e conhecimentos repassados. Por serem exemplos na pesquisa, educação e como pessoas possuidoras de valores imprescindíveis ao ser humano.

À professora Líliliana Madalena Gramani pelos seus esforços ao longo do curso, orientações e cobranças que atuaram de forma motivadora nos impulsionando para frente.

Ao professor Anselmo Chaves Neto, pelos ensinamentos, incentivos e momentos de descontração nesta árdua caminhada.

Aos professores Arinei Carlos Lindbeck da Silva, Neida Maria Patias Volpi e Cassius Tadeu Scarpin pelos ensinamentos.

À minha família, especialmente à minha esposa Lúcia Celina, aos meus filhos Vitória, João Levi e Raquel por entenderem a minha ausência durante o curso.

À Universidade Federal do Paraná, Universidade Estadual do Oeste do Paraná, Itaipu, CEASB e todas as pessoas que contribuíram para a viabilização do doutorado.

Aos amigos e colegas do curso, especialmente aqueles que não faltaram com uma palavra amiga nos momentos difíceis.

À Universidade Tecnológica Federal do Paraná, a qual sou filiado. Aos integrantes do Departamento de Matemática. Ao meu amigo Samuel Bellido Rodrigues pelas sugestões, incentivos e relevantes contribuições.

RESUMO

Nesta tese de doutorado, é proposto um novo método híbrido *wavelet*-neural integrado com um amostrador *bootstrap* para projeção pontual e intervalar de séries temporais estocásticas. Basicamente, combinam-se os métodos de encolhimento e de decomposição *Wavelet* no pré-processamento dos dados e, em seguida, uma Rede Neural Artificial (RNA) é usada para produzir as de previsões pontuais. A medida de incerteza do modelo RNA é obtida com a utilização de uma amostragem *bootstrap* dos resíduos do modelo RNA ajustado à série temporal subjacente. A fim de se obter o intervalo de confiança, calculou-se a média ponderada das previsões de B séries temporais oriundas do processo *bootstrap*, sendo os pesos determinados via otimização de um problema de programação não linear cuja função objetivo é a minimização da raiz quadrada do erro quadrático médio entre a combinação linear das B previsões e a série temporal subjacente. Além do intervalo de confiança, obtém-se também a estimativa do intervalo de previsão, sendo este último mais amplo que o primeiro, pois engloba as variâncias do modelo de regressão RNA e dos ruídos. A estimativa do desvio padrão dos ruídos foi alcançada com o treinamento de uma RNA com função custo oriunda da função de máxima verossimilhança log-normal, otimizada por meio da meta-heurística PSO. Para se averiguar a eficiência do método preditivo proposto, foram realizados experimentos computacionais para previsões pontuais envolvendo séries temporais utilizadas com frequência para este fim, podendo ser encontradas com facilidade em publicações da literatura especializada. As séries temporais referidas são as seguintes: *Canadian Lynx*, *Wolf's Sunspot* e *Exchange Rate*. Os desempenhos preditivos alcançados pelo método proposto, em relação às abordagens de outros autores, são efetivos e consideráveis. Em particular, os intervalos de confiança e previsão foram estimados para uma série temporal de vazão média mensal afluyente da hidrelétrica de Itaipu, em Foz do Iguaçu, Brasil. Neste caso, para efeito de comparação, usaram-se os seguintes métodos de previsão: RNA, Box & Jenkins e decomposição *wavelet* integrada com rede neural artificial e ARIMA (*Wavelet*-RNA e *Wavelet*-ARIMA). Comparações com resultados de previsões obtidos através de modelos de Box & Jenkins e RNA, quando usados individualmente, constata consideráveis ganhos preditivos auferidos com o uso do método proposto, reduzindo o erro preditivo em 62%, aproximadamente. Na comparação com os métodos compostos *Wavelet*-RNA e *Wavelet*-ARIMA, a redução do erro foi da ordem de 54%.

Palavras-chave: Séries temporais, *Wavelet*, Redes Neurais Artificiais, Programação matemática, Amostrador *Bootstrap*.

ABSTRACT

In this doctoral thesis, it's proposed a new hybrid integrated *wavelet*-neural method with a *bootstrap* sampler for point and interval projection of stochastic time series. Basically, the methods *Wavelet* shrinkage and decomposition are combined in the pre-processing of data and then an Artificial Neural Network (ANN) is used to produce the point predictions. The ANN template uncertainty measurement is achieved with the use of a *bootstrap* sample of ANN template's waste adjusted to the underlying time series. In order to obtain the confidence interval, it was calculated the weighted average forecast of B time series derived from the *bootstrap* process, and these weights were determined by optimization of a nonlinear programming problem whose objective function is the minimization of square root of the medium squared error between the linear combination of N forecasts and the underlying time series. Beyond the confidence interval it's also obtained the estimate of prediction interval, the latter is wider than the first, because it includes the variances ANN regression model and the noises. The estimate standard deviation of the noises was achieved with the training of an ANN with cost function derived of the lognormal likelihood maximum function optimized by PSO meta-heuristics. To ascertain the efficiency of the predictive method proposed, computational experiments were performed for point predictions involving time series frequently used for this purpose and may be easily found in specialized literature publications. The time series mentioned are: Canadian Lynx, Wolf's Sunspot and Exchange Rate. The predictive performances achieved by the proposed method, in relation to approaches from other authors are effective and substantial. In particular, confidence and prediction intervals were estimated for a time series of monthly average flow tributary of Itaipu dam, in Foz do Iguaçu, Brazil. In this case, for comparison, they used the following prediction methods: ANN, Box & Jenkins, *wavelet* decomposition integrated with artificial neural networks and ARIMA (*Wavelet*-ANN and *Wavelet*-ARIMA). Comparison with results of predictions obtained through model Box and Jenkins and ANN, when used alone, find considerable predictive gains obtained using the proposed method, reducing the predictive error in 62% approximately. In comparison to the compounds methods *Wavelet*-ANN and *Wavelet*-ARIMA the reducing of error was approximately 54%.

Keywords: Time series, *Wavelet*, Artificial Neural Networks, Mathematical programming, *Bootstrap* Sampler.

LISTA DE FIGURAS

FIGURA 1 - FLUXOGRAMA COM AS ETAPAS DO MÉTODO PROPOSTO	18
FIGURA 2 - GRÁFICOS DAS FUNÇÕES ESCALA E WAVELET DE HAAR.....	22
FIGURA 3 - GRÁFICOS DAS FUNÇÕES LIMIAR DURO E SUAVE.	25
FIGURA 4 - RNA MULTILAYER FEEDFORWARD.....	27
FIGURA 5 - RNA <i>FEEDFORWARD</i> COM UMA CAMADA OCULTA.	27
FIGURA 6 - MONTAGEM DO CONJUNTO DE TREINAMENTO.	30
FIGURA 7 - GRÁFICO DA SÉRIE TEMPORAL <i>CANADIAN LYNX</i> NO PERÍODO 1821-1934	43
FIGURA 8 - GRÁFICO DA SÉRIE TEMPORAL <i>WOLF'S SUNSPOT</i> NO PERÍODO 1700-1987	43
FIGURA 9 - GRÁFICO DA SÉRIE TEMPORAL <i>EXCHANGE RATE</i> COM 731 OBSERVAÇÕES (1980-1993)	44
FIGURA 10 - SÉRIE DE VAZÕES MENSIS OBTIDAS NO POSTO 266-ITAIPU NO PERÍODO 1970-2010.....	45
FIGURA 11 - PERIODOGRAMA ACUMULADO	47
FIGURA 12 - FLUXOGRAMA COM AS ETAPAS DO MÉTODO <i>WAVELET_ARIMA</i>	49
FIGURA 13 - GRÁFICO DE SINAIS SEM E COM RUÍDO, FILTRADO E COMPONENTE <i>WAVELET</i>	52
FIGURA 14 - FUNÇÃO <i>WAVELET</i> DB4 E COIF4.....	55
FIGURA 15 - FLUXOGRAMA COM ETAPAS DO MÉTODO PROPOSTO: PREVISÕES PONTUAIS.	58
FIGURA 16 - FLUXOGRAMA DA REDE-PSO QUE ESTIMA A VARIÂNCIA DO RUÍDO	61
FIGURA 17 - GRÁFICOS DAS SÉRIES TENDÊNCIA E RESÍDUOS ORIUNDOS DO EW DA LOG <i>CANADIAN LYNX</i>	63
FIGURA 18 - GRÁFICOS DAS SÉRIES LOG <i>CANADIAN LYNX</i> E PREVISÕES PELO MÉTODO PROPOSTO – AMOSTRA DE TESTE.....	64
FIGURA 19 - GRÁFICOS DAS SÉRIES: TENDÊNCIA, <i>X</i> E <i>Z</i> - <i>WOLF'S SUNSPOT</i>	66
FIGURA 20 - GRÁFICOS DAS SÉRIES: RESÍDUOS, <i>X</i> E <i>Z</i> - <i>WOLF'S SUNSPOT</i> . 66	

FIGURA 21 - GRÁFICOS DAS SÉRIES WOLF'S SUNSPOT E PREVISÕES PELO MÉTODO PROPOSTO – AMOSTRA DE TESTE.	67
FIGURA 22 - GRÁFICOS DAS SÉRIES LOG <i>EXCHANGE RATE</i> , TENDÊNCIA E RESÍDUOS.....	68
FIGURA 23 - GRÁFICOS DAS SÉRIES LOG <i>EXCHANGE RATE</i> E PREVISÕES PELO MÉTODO PROPOSTO – AMOSTRA DE TESTE.....	70
FIGURA 24 - CORRELOGRAMA DA SÉRIE DE VAZÕES MENSAS.....	75
FIGURA 25 - AUTOCORRELAÇÕES DOS RESÍDUOS: ARIMA(1,0,2)X(0,1,2)12...76	
FIGURA 26 - PERIODOGRAMA PARA OS RESÍDUOS: ARIMA(1,2,2)X(0,1,2)12 ..77	
FIGURA 27 - DIAGRAMAS DE DISPERSÃO ENTRE A AMOSTRA DE TESTE E PREVISÕES POR RNA E ARIMA.....	78
FIGURA 28 - GRÁFICO DA SÉRIE DE VAZÕES E COMPONENTES <i>WAVELET</i> ...78	
FIGURA 29 - GRÁFICOS DAS SÉRIES DE VAZÕES, TENDÊNCIA E RESÍDUOS	80
FIGURA 30 - GRÁFICO DOS RESÍDUOS EW	81
FIGURA 31 - FAC DA SÉRIE DE RESÍDUOS ORIUNDOS DO EW.....	81
FIGURA 32 - ESCALOGRAMA – COEFICIENTES DE DETALHES NA DW DA SÉRIE TENDÊNCIA.....	83
FIGURA 33 - GRÁFICOS DAS COMPONENTES <i>WAVELET</i> DE DETALHES– SÉRIE TENDÊNCIAS.....	83
FIGURA 34 - GRÁFICOS DAS SÉRIES TEMPORAIS X E Z – TENDÊNCIA.....	84
FIGURA 35 - GRÁFICOS DAS SÉRIES TENDÊNCIA, X E Z . $TENDÊNCIA \cong X + Z$	84
FIGURA 36 - DIAGRAMA DE DISPERSÃO ENTRE A AMOSTRA DE TESTE E PREVISÕES PELO MÉTODO PROPOSTO.	86
FIGURA 37 - FLUXOGRAMA COM AS ETAPAS DO PROCEDIMENTO <i>BOOTSTRAP</i> NA ESTIMATIVA DO IC.	87
FIGURA 38 - IC PREVISÕES DA SÉRIE X : TENDÊNCIA.....	88
FIGURA 39 - IC PREVISÕES DA SÉRIE Z : TENDÊNCIA.....	88
FIGURA 40 - IC PREVISÕES DA SÉRIE DE VAZÕES – AMOSTRA DE TESTE	89
FIGURA 41 - COMPORTAMENTO DA FUNÇÃO C DURANTE O PROCESSO DE OTIMIZAÇÃO	90
FIGURA 42 - INTERVALO DE PREVISÃO PARA A SÉRIE DE VAZÕES – AMOSTRA DE TESTE	91

FIGURA 43 - GRÁFICOS DAS VAZÕES OBSERVADAS (AMOSTRA DE TESTE) E SUAS PREVISÕES PELO MÉTODO PROPOSTO	91
FIGURA 44 - MAPE AMOSTRA DE TESTE PARA OS MÉTODOS <i>WAVELET_RNA</i> , <i>WAVELET_ARIMA</i> E PROPOSTO.	92
FIGURA 45 - APE NA AMOSTRA DE TESTE PARA OS MÉTODOS <i>WAVELET_RNA</i> E PROPOSTO.	93

LISTA DE QUADROS

QUADRO 1 - ALGUMAS CARACTERÍSTICAS DAS FAMÍLIAS DB, SYM E COIF.	55
QUADRO 2 - CONFIGURAÇÕES RNA E ORDENS DOS MODELOS AR E MA PARA A2, D1 E D2.....	79
QUADRO 3 - MAPE E CONFIGURAÇÕES DAS RNA USADAS NA MODELAGEM DAS SÉRIES X E Z PROVENIENTES DAS SÉRIES TENDÊNCIA E RESÍDUOS EW.....	85

LISTA DE TABELAS

TABELA 1 - DESCRIÇÃO DAS SÉRIES <i>CANADIAN LYNX</i> , <i>WOLF'S SUNSPOT</i> E <i>EXCHANGE RATE</i> E TAMANHO DAS AMOSTRAS UTILIZADAS NA MODELAGEM.....	46
TABELA 2 - BASES <i>WAVELET</i> UTILIZADAS NO ENCOLHIMENTO DA SÉRIE <i>CANADIAN LYNX</i> E ERROS DE PREVISÕES DA AMOSTRA DE TESTE.....	62
TABELA 3 - PREVISÕES PARA AS 10 ÚLTIMAS OBSERVAÇÕES DA AMOSTRA DE TESTE - <i>CANADIAN LYNX</i>	63
TABELA 4 - COMPARATIVO MSE E MAE ENTRE MÉTODOS DE PREVISÃO APLICADOS À SÉRIE <i>CANADIAN LYNX</i>	64
TABELA 5 - MAPE E MAE PARA MODELAGENS ORIUNDAS DE ENCOLHIMENTOS <i>WAVELET</i> DISTINTOS- <i>WOLF'S SUNSPOT</i> ..	65
TABELA 6 - PREVISÕES PARA AS 10 ÚLTIMAS OBSERVAÇÕES DA AMOSTRA DE TESTE - <i>WOLF'S SUNSPOT</i>	67
TABELA 7 - COMPARATIVO MSE E MAE ENTRE MÉTODOS DE PREVISÃO APLICADOS À SÉRIE <i>WOLF'S SUNSPOT</i>	68
TABELA 8 - PREVISÕES PARA AS 10 ÚLTIMAS OBSERVAÇÕES DA AMOSTRA DE TESTE - <i>EXCHANGE RATE</i>	69
TABELA 9 - COMPARATIVO MSE E MAE ENTRE MÉTODOS DE PREVISÃO APLICADOS À SÉRIE LOG <i>EXCHANGE RATE</i>	69
TABELA 10 - APE MÍNIMO, MÁXIMO E MAPE PARA SEIS ESTRUTURAS DISTINTAS DE RNA-MLP.....	77
TABELA 11 - MAPE E MAE PARA MODELAGENS ORIUNDAS DE ENCOLHIMENTOS <i>WAVELET</i> DISTINTOS- SÉRIE DE VAZÕES.....	79
TABELA 12 - TESTE DE INDEPENDÊNCIA ESTATÍSTICA BDS.....	82
TABELA 13 - PREVISÕES PARA AS 12 ÚLTIMAS OBSERVAÇÕES DA AMOSTRA DE TESTE – VAZÕES MENS AIS POSTO 266-ITAIPU.....	85
TABELA 14 - PARÂMETROS META-HEURÍSTICA PSO.....	90
TABELA 15 - MAPE E MAE PARA AMOSTRA DE TESTE RETIRADA DA SÉRIE DE VAZÕES POSTO 266-ITAIPU.....	92

SUMÁRIO

1 INTRODUÇÃO	14
1.1 OBJETIVOS	17
1.1.1 Objetivo Geral	17
1.1.2 Objetivos Específicos	17
1.2 JUSTIFICATIVA	17
1.3 ESTRUTURA DO TRABALHO	20
2 REVISÃO DA LITERATURA	21
2.1 TEORIA <i>WAVELET</i>	21
2.2 ENCOLHIMENTO <i>WAVELET</i>	24
2.3 REDES NEURAIS ARTIFICIAIS	26
2.4 COMBINAÇÃO DE PREVISÕES	30
2.4.1 Otimização	32
2.5 <i>BOOTSTRAP</i>	34
2.6 INTERVALO DE CONFIANÇA E PREVISÃO	36
2.7 META-HEURÍSTICA PSO	39
3 MATERIAIS E MÉTODOS	42
3.1 MATERIAS	42
3.1.1 Série temporal <i>Canadian Lynx</i> - Experimento Computacional 1	42
3.1.2 Série Temporal <i>Wolf's Sunspot</i> – Experimento Computacional 1	43
3.1.3 Série temporal <i>Exchange Rate</i> – Experimento Computacional 1	43
3.1.4 Série de Vazões Mensais – Experimento Computacional 2	44
3.2 MÉTODOS	45
3.2.1 Experimento Computacional 1	45
3.2.2 ARIMA	46
3.2.3 Redes Neurais Artificiais <i>Feedforward</i>	48
3.2.4 Wavelet-ARIMA e Wavelet-RNA	48
4 MÉTODO PROPOSTO	51
4.1 ENCOLHIMENTO <i>WAVELET</i> DA SÉRIE OBSERVADA	51
4.2 DECOMPOSIÇÃO <i>WAVELET</i> : TENDÊNCIA E RESÍDUO	53
4.3 PREVISÕES PONTUAIS	57
4.4 PROPOSTA DE INTERVALO DE CONFIANÇA	58

4.5 PREVISÕES INTERVALARES	60
5 EXPERIMENTO COMPUTACIONAL 1: PREVISÕES PONTUAIS	62
5.1 SÉRIE TEMPORAL CANADIAN LYNX	62
5.2 SÉRIE TEMPORAL <i>WOLF'S SUNSPOT</i>	65
5.3 SÉRIE TEMPORAL <i>EXCHANGE RATE</i>	68
6 EXPERIMENTO COMPUTACIONAL 2: PREVISÕES INTERVALARES	71
6.1 O PROBLEMA DA PREVISÃO DE VAZÕES	71
6.2 MÉTODOS: ARIMA, RNA, <i>WAVELET</i> -ARIMA E <i>WAVELET</i> -RNA.....	74
6.3 MÉTODO PROPOSTO: RESULTADOS EMPÍRICOS	79
6.3.1 Filtragem <i>Wavelet</i>	80
6.3.2 Decomposição <i>Wavelet</i>	82
6.3.3 Modelagem por RNA: previsões pontuais	84
6.3.4 Intervalo de confiança por <i>Bootstrap</i>	86
6.3.5 Previsões Intervalares	89
6.4 COMPARATIVO ENTRE OS MÉTODOS.....	92
7 CONCLUSÕES	94
7.1 SUGESTÕES PARA TRABALHOS FUTUROS	96
REFERÊNCIAS.....	98

1 INTRODUÇÃO

Desde 1927 quando George Udny Yule lançou a ideia que série temporal pode ser vista como sendo gerada por um conjunto de variáveis aleatórias conjuntamente distribuídas e aplicou modelos *autorregressivos* (ARs), muitos foram os avanços na tentativa de melhorar a acurácia na modelagem de séries temporais.

Box e Jenkins (1970) apresentaram metodologia que fazia a integração de conhecimentos existentes na época. A metodologia consiste em ajustar modelos *autorregressivos integrados de médias móveis* (ARIMA) a um conjunto de dados a partir de três estágios: identificação do modelo, estimação dos parâmetros e verificação do modelo ajustado.

A metodologia de Box e Jenkins é largamente aplicada, mas recomendada para séries temporais provenientes de processos lineares. Para séries temporais oriundas de processos não lineares são indicados outros métodos, sendo predominante para este caso o uso de Redes Neurais Artificiais (RNA). Observado que as RNAs podem ser utilizadas tanto na abordagem linear quanto na não linear. Ballini (2000); Wang *et al.*(2006); Valenzuela *et al.* (2008); Ferreira (2012); Abreu *et al.*(2012); Teixeira, Jr *et al.* (2012a), entre outros, utilizaram Redes Neurais Artificiais (RNA) na previsão de séries temporais.

Revisando 20 anos (1985-2005) de pesquisa sobre previsão de séries temporais, Gooijer e Hyndman (2006) concluíram que importantes progressos foram alcançados. Os autores salientam que numerosos estudos atestam o sucesso das redes neurais artificiais na previsão de séries temporais e apontam temas que necessitam de um maior desenvolvimento, entre eles, a combinação de previsões, onde ainda não há uma resposta para a melhor estratégia de combinação de métodos lineares e não lineares.

Zou e Yang (2004); Deutsch *et al.* (1994); Martins (2011); Fiordaliso (1998); Fang (2003); Cavalieri e Ribeiro (2011); Bunn (1988) são autores que desenvolveram trabalhos que mostram empiricamente a eficácia da combinação de previsões. Em sua pesquisa clássica, Clemen (1989) concluiu que combinações de previsões têm se mostrado métodos práticos e úteis. Teixeira, Siqueira e Teixeira Jr. (2015) alcançaram resultados que fortalecem a conclusão que a combinação de

métodos preditivos pode auferir ganhos preditivos consideráveis em relação a outros que não o fazem.

Outra técnica que vem sendo usada com sucesso nas previsões de séries temporais é a abordagem *wavelet*. Embora os primeiros estudos no campo das *wavelets* tenham ocorridos em 1909 (HAAR, 1910), somente em 1985, Stephane Mallat deu às *wavelets* um grande impulso através do seu trabalho em processamento digital de imagens, construindo a primeira *wavelet* não-trivial (suave) (LIMA, 2002). Daubechies (1988) construiu um conjunto de bases ortonormais de *wavelet* suaves, passando a ser uma das principais referências para as atuais aplicações de *wavelet*.

Lima (2011) utilizou *wavelet* no seu trabalho de previsão de séries econômicas, separando a parte determinística dos ruídos. Teixeira Jr *et al.* (2012a) optaram por um método de combinação em que as componentes *wavelet* de uma série temporal constituem os padrões de entrada de uma RNA *feedforward* MLP (*multi-layer perceptron*), cuja saída fornece a previsão da série temporal. Karthikeyan e Kumar (2013); Tiwari e Chatterjee (2010); Kisi (2010); Kisi e Cimen (2011); Nalley *et al.* (2012) também utilizaram *wavelet* em seus trabalhos de previsão de séries temporais e obtiveram ganhos preditivos relevantes.

Se a série temporal $f(.) := (f(1), f(2), \dots, f(T))$ for fortemente ruidosa, é recomendada a sua filtragem para então modelá-la. Note que $f(.)$ é tal que, para cada instante t , pode ser teoricamente decomposta como: $f(t) = \tilde{f}(t) + e_t$, onde: (i) e_t é um estado (ou realização) de uma variável aleatória independente e identicamente distribuída (iid), com média zero e variância constante; e (ii) $\tilde{f}(t)$ é a componente determinística (que, de fato, é passível de ser prevista pontualmente ao contrário de e_t) de $f(t)$.

Entre as técnicas de filtragem de séries temporais existentes, destaca-se o encolhimento (*shrinkage*) *wavelet*, sendo David L. Donoho e Iain M. Johnstone os precursores nesta área com os trabalhos: Donoho e Johnstone (1994, 1995, 1998), entre outros. A filtragem *wavelet* pode ser abordada com uma regressão não paramétrica, onde a função $f(x_i)$ no modelo $y_i = f(x_i) + \varepsilon_i$ é obtida dos dados ruidosos y_i , sem que seja assumida uma estrutura paramétrica para $f(x_i)$ (ABRAMOVICH *et al.* 2000). Uma das soluções para esta abordagem é a expansão *wavelet* do sinal e eliminação, manutenção ou atenuação dos coeficientes das

componentes de detalhes mediante regras de limiarização. Outra técnica de filtragem de sinal bastante utilizada é a *análise espectral singular* (SSA – *singular spectrum analysis*), que se trata de uma abordagem relativamente nova que incorpora elementos de análise clássica de séries temporais, estatística multivariada, geometria multivariada, sistemas dinâmicos e processamentos de sinais (ELSNER e TSONIS, 1996). A partir do método SSA, uma série temporal pode ser decomposta ortogonalmente em componentes que podem ser agrupadas em três grupos distintos: tendência, harmônica e ruído, sendo possível a transformação da série temporal observada em uma série menos ruidosa.

Modelos matemáticos e o avanço das tecnologias da informação estão propiciando cada vez mais o uso de previsões, sendo utilizadas por profissionais das engenharias, ciências e negócios. Tomadores de decisão empregam previsões em análises e planejamento de necessidades operacionais futuras em áreas diversas, tais como: precipitação atmosférica, índices da bolsa de valores, planejamento energético, entre outras.

Wang *et al.* (2006a) utilizaram redes neurais artificiais para fazer a previsão de vazões na cabeceira do rio Amarelo na China. Maceira *et al.* (1999) fizeram previsões de vazões em usinas hidrelétricas do Brasil através de modelos ARIMA. Lima (2011) utilizou filtros de Kalman, *wavelet* e redes neurais para a previsão de séries temporais financeiras. Yoon *et al.* (2011) fizeram um estudo comparativo de redes neurais artificiais e máquinas de vetores de suporte para a previsão de níveis de águas subterrâneas em um aquífero costeiro. Martins e Werner (2014) trabalharam com previsão de séries industriais a partir da combinação de previsões obtidas através de modelos ARIMA e redes neurais artificiais (RNA). Teixeira *et al.* (2012a) combinaram RNA e decomposição *wavelet* na previsão de radiação solar direta.

Diante do exposto nos parágrafos anteriores, percebe-se uma grande diversidade de métodos e aplicações das séries temporais, sendo a melhoria da acurácia das previsões o cerne dos estudos encontrados nesta área. O presente trabalho está inserido neste contexto e se propõe a apresentar um método de previsão para séries temporais envolvendo *wavelet* na filtragem das séries temporais e na decomposição ortogonal do resultado da filtragem, modelagem neural e *bootstrap* na estimativa da incerteza que envolve o processo de geração das previsões.

1.1 OBJETIVOS

1.1.1 Objetivo Geral

Analisar séries temporais e propor um método de previsão que contemple tanto os sinais com estruturas de autodependência linear como não linear.

1.1.2 Objetivos Específicos

- Avaliar a importância da filtragem da série para a qualidade das previsões;
- Mostrar empiricamente a eficiência da decomposição *wavelet* na previsão de séries temporais;
- Estimar as incertezas no processo das previsões realizadas no presente trabalho;
- Comparar os resultados alcançados com o método proposto aos obtidos por métodos de previsão tradicionais (ARIMA e RNA).

1.2 JUSTIFICATIVA

Segundo Levine *et al.* (2000), uma série temporal é um conjunto de dados numéricos obtidos durante períodos regulares ao longo do tempo. O pressuposto básico da análise de séries temporais é de que fatores que influenciaram padrões da atividade no passado e no presente continuarão a fazê-lo, da mesma maneira, no futuro. Portanto, os principais objetivos da análise de séries temporais são identificar e isolar esses fatores de influência, para fins de previsão (prognósticos), bem como para planejamento e controle gerencial. Para alcançar esses objetivos foram criados muitos modelos matemáticos para pesquisar as flutuações entre os fatores que compõem uma série temporal. Desta forma, a principal motivação para este trabalho é o desenvolvimento de um método que apresente vantagens preditivas quando comparado com métodos tradicionais em previsões de séries em geral, inclusive aquelas que apresentam características de não linearidade, sazonalidade e não estacionariedade. Para tanto, o método proposto possui as seguintes etapas:

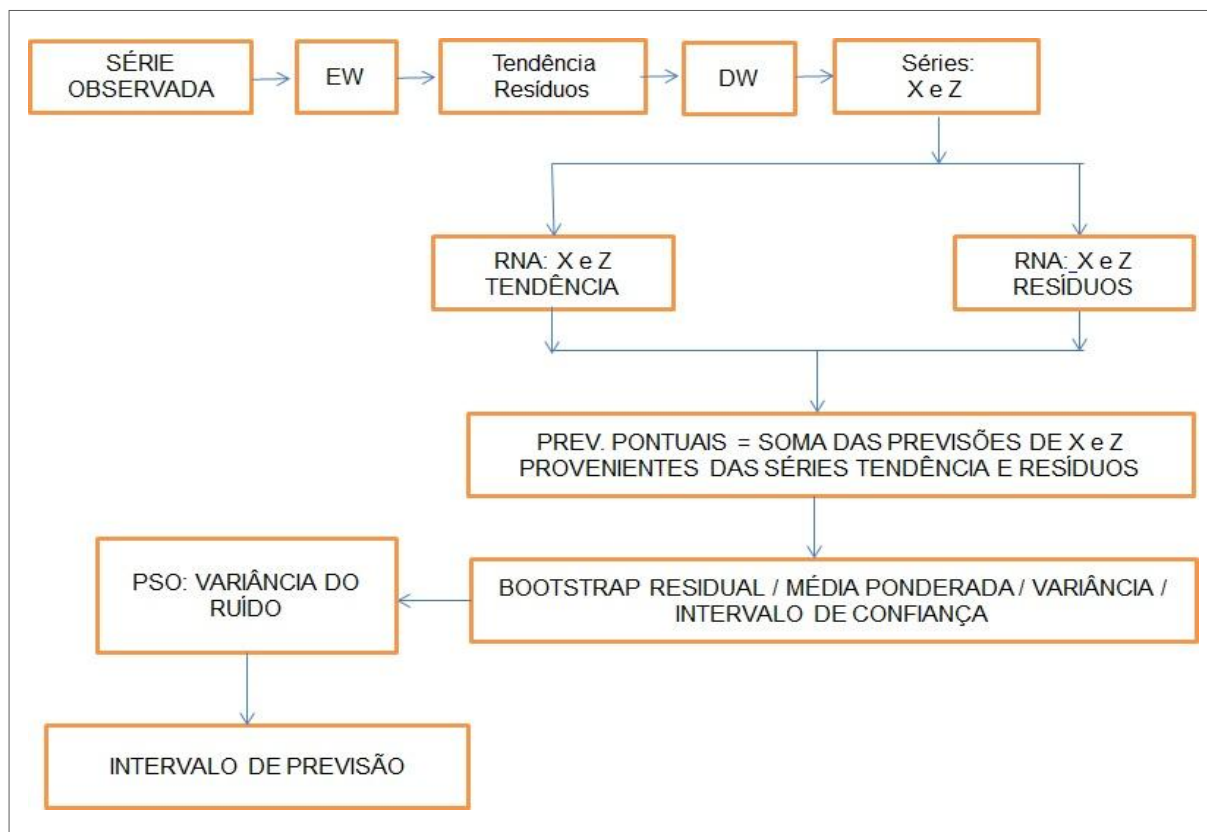
- (i) Alisamento da série temporal por encolhimento *wavelet* (EW), obtendo-se uma série temporal menos ruidosa (Tendência) e outra constituída pelos

resíduos *wavelet*. A soma dessas duas séries temporais é igual à observada;

- (ii) Decomposição ortogonal (usando *wavelet*) das séries temporais Tendência e resíduos *wavelet*. Da decomposição *wavelet* (DW) resultam componentes de detalhes e aproximação, que são agrupadas em duas séries denominadas por X e Z;
- (iii) Modelagem das séries temporais X e Z oriundas da decomposição *wavelet* das séries temporais Tendência e resíduos por RNAs. As previsões pontuais da série temporal observada resultam da soma das previsões das séries X e Z;
- (iv) Amostras *bootstrap residual* permitem estimar as incertezas do modelo neural e consequente intervalo de confiança; e
- (v) Estimativa da variância dos resíduos de previsão a partir de RNA associada à meta-heurística PSO, permite o estabelecimento de intervalo de previsão.

O fluxograma da Figura 1 ilustra as etapas descritas.

FIGURA 1 - FLUXOGRAMA COM AS ETAPAS DO MÉTODO PROPOSTO



FONTE: O autor (2015).

A previsão de séries temporais ruidosas envolve duas abordagens básicas. Na primeira, ignoram-se os ruídos e modela-se a série na expectativa de se capturar características determinísticas da série temporal. Na segunda, faz-se a filtragem da série procurando identificar as componentes de ruídos e determinística. De maneira geral, a segunda abordagem é mais eficiente que a primeira (HASSANI; ZHIGLJAVSKY, 2009). Menezes *et al.* (2014) modelaram séries temporais através de vários métodos preditivos, comparando os resultados de previsões oriundas das séries observadas (dados não tratados) com os resultados obtidos a partir das séries filtradas. Concluíram que a filtragem de séries temporais antes da modelagem apresenta ganho significativo nas previsões.

Por sua vez, da Análise *Wavelet*, advém importantes métodos auxiliares de pré-processamento (decomposição, filtragem ou alisamento) de uma série de tempo, antes de sua efetiva modelagem com um método preditivo base plausível, de forma a gerar um novo método para se prever séries temporais (AQUINO *et al.* 2009). Desde que sejam utilizados de forma adequada com os métodos preditivos, é possível se obter ganhos preditivos relevantes. Pode-se verificar este fato, dentre muitas outras referências, em: Donoho e Jonhstone (1994); Donoho *et al.* (1995); Lei e Ran (2008).

Embora a análise de Fourier seja uma das formas mais tradicionais para o tratamento de sinais, as transformadas *wavelets* são mais completas, fornecendo uma melhor descrição do sinal. Karthikeyan e Kumar (2013); Tiwari e Chatterjee (2010); Kisi (2010); Kisi e Cimen (2011); Nalley *et al.* (2012) utilizaram decomposição *wavelet* em seus trabalhos de previsão de séries temporais e concluíram o bom desempenho desta técnica. Teixeira Jr *et al.* (2012a), apresentaram um método em que as componentes *wavelets* de uma série temporal constituem os padrões de entrada de uma RNA *feedforward* MLP, cuja saída fornece a previsão da série temporal. Com esta metodologia, chegaram a resultados melhores do que os obtidos por RNA convencional (sem considerar componentes *wavelet* como padrões de entrada).

Combinar linearmente ou não linearmente previsões provenientes de métodos distintos, de maneira geral, leva a ganhos preditivos consideráveis. Hamilton (1994); Newbold e Granger (1974); Winkler e Makridakis (1982); Teixeira Jr (2013) apontam que a combinação de previsões implica na diminuição do erro preditivo. Segundo

Werner e Ribeiro (2006); Wallis (2011) não há ainda uma combinação de métodos preditivos que seja considerada a melhor na projeção de séries temporais. A combinação de previsores e previsões aliada à filtragem da série temporal e decomposição *wavelet* se apresenta como uma área de pesquisa aberta a discussões e consequente estabelecimento de novos métodos que venham oferecer ganho na análise e previsão de séries temporais.

1.3 ESTRUTURA DO TRABALHO

Esta tese está dividida em sete capítulos: (1) Introdução; (2) Revisão Bibliográfica; (3) Materiais e Métodos; (4) Método Proposto; (5) Experimento Computacional 1; (6) Experimento Computacional 2; (7) Conclusões. No capítulo 2 encontram-se conceitos e descrições de técnicas a serem usadas na pesquisa, tais como: *Wavelet*, RNA-MLP, *bootstrap* e PSO. No capítulo 3 está a descrição dos dados e métodos empregados. O Capítulo 4 traz os detalhes do método proposto nesta tese. Já o capítulo 5 traz resultados alcançados com a utilização do método proposto na previsão das séries temporais *Canadian Lynx*, *Wolf's Sunspot* e *Exchange Rate*, frequentemente utilizadas na literatura especializada. No capítulo 6 foi modelada uma série de vazões mensais coletadas no posto 266-Itaipu, utilizando-se métodos individuais e compostos. Além, da geração de intervalos de confiança e previsão por meio do método proposto. Finalmente o capítulo 7 traz as principais conclusões desta pesquisa e algumas sugestões de trabalhos futuros.

2 REVISÃO DA LITERATURA

Neste capítulo é feita uma sucinta revisão acerca das abordagens utilizadas na construção da metodologia híbrida proposta, são elas: *Wavelet*, RNA; combinação de previsões; e amostrador *Bootstrap*.

2.1 TEORIA WAVELET

Uma função *wavelet* é uma forma de onda que, ao contrário das senóides, podem apresentar (dependendo da família) forma irregular e assimétrica. Enquanto na análise de Fourier o sinal é decomposto em ondas senoidais de várias bandas de frequências; na análise *wavelet*, um sinal $f(x)$ é decomposto a partir de informações de escala (compressão e dilatação) e translação de uma função *wavelet* (MORETTIN, 1999).

O mapeamento das componentes de alta frequência de um sinal $f(x)$ é realizado pela função *wavelet* $\omega(x)$ e as de baixa frequência descritas pela função escala $\phi(x)$. Entre as famílias de funções *wavelet* vigentes na literatura, a mais antiga é a de Haar, a qual é definida na equação (1).

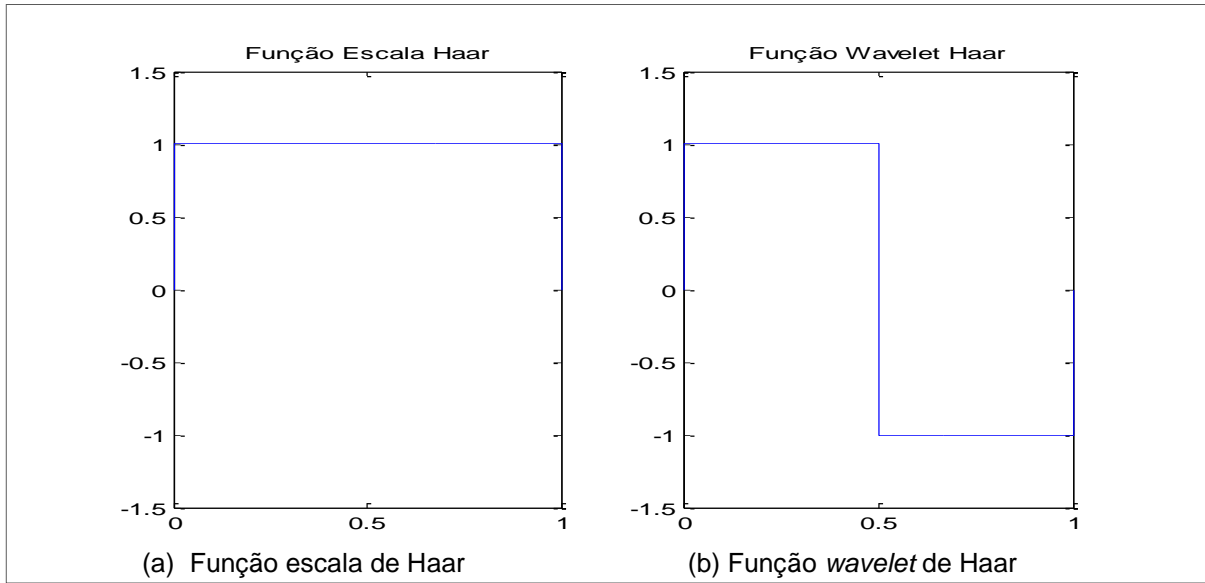
$$\omega(x) = \begin{cases} 1, & \text{se } x \in [0, 1/2) \\ -1, & \text{se } x \in [1/2, 1) \\ 0, & \text{caso contrário} \end{cases} \quad (1)$$

Na equação (2), por sua vez, encontra-se a definição da função escala da família de Haar.

$$\phi(x) = \begin{cases} 1, & \text{se } x \in [0, 1) \\ 0, & \text{caso contrário} \end{cases} \quad (2)$$

Nas Figuras 2a e 2b estão representados os gráficos das funções escala e *wavelet* de Haar, respectivamente.

FIGURA 2 - GRÁFICOS DAS FUNÇÕES ESCALA E WAVELET DE HAAR.



FONTE: O autor (2015).

Do ponto de vista matemático, as funções *wavelet* são elementos do espaço de Hilbert l^2 formado por funções quadraticamente somáveis. A função $\omega(\cdot)$ pertencente a l^2 é chamada de função *wavelet* se, e somente se, as funções $\omega_{m,n}(\cdot) = 2^{\frac{m}{2}} \omega(2^m(\cdot) - n)$ onde $m, n \in \mathbb{Z}$, formam uma base ortonormal para o espaço de Hilbert l^2 , onde, segundo Levan e Kubrusly (2003), o parâmetro m é chamado de parâmetro de escala e n , de parâmetro de translação. Desta forma, a função $f(\cdot) \in l^2$ admite sua expansão ortogonal por meio de uma série de Fourier, em termos de uma base ortonormal *wavelet* $\{\omega_{m,n}(\cdot)\}_{(m,n) \in \mathbb{Z} \times \mathbb{Z}}$ para l^2 , definida como na equação (3).

$$f(\cdot) = \sum_{m \in \mathbb{Z}} \sum_{n \in \mathbb{Z}} \langle f(\cdot), \omega_{m,n}(\cdot) \rangle \omega_{m,n}(\cdot). \quad (3)$$

De acordo com Kubrusly e Levan (2006) o subespaço fechado, equação (4),

$$W_m(\omega) := \left(\text{span}\{\omega_{m,n}(\cdot)\}_{n \in \mathbb{Z}} \right)^-, \quad (4)$$

pertencente a l^2 é chamado de subespaço de detalhes, na escala m . Segundo Levan e Kubrusly (2003), a projeção ortogonal de $f(\cdot)$, sobre o subespaço de detalhes $W_m(\omega)$, é definida pela soma parcial definida na equação (5).

$$f_{W_m(\omega)}(\cdot) := \sum_{n \in \mathbb{Z}} \langle f(\cdot), \omega_{m,n}(\cdot) \rangle \omega_{m,n}(\cdot). \quad (5)$$

Segundo Daubechies (1988), a projeção $f_{W_m(\omega)}(\cdot)$ pode ser referida como uma componente de detalhe *wavelet* de $f(\cdot)$, na escala m , sobre $W_m(\omega)$.

A função $\phi(\cdot) \in l^2$ é chamada de função (*wavelet*) *escala* se, e somente se, a coleção de funções $\phi_{m,n}(\cdot) = 2^{\frac{m}{2}} \phi(2^m(\cdot) - n)$, onde $m, n \in \mathbb{Z}$, são tais que $\langle \phi_{l,i}(\cdot), \phi_{j,k}(\cdot) \rangle = 0$ sempre que $l = j$ e $i \neq k$, e $\langle \phi_{l,i}(\cdot), \phi_{j,k}(\cdot) \rangle \neq 0$, caso contrário.

De acordo com Mallat (1998), o subespaço fechado $V_m(\phi)$ de l^2 , onde $V_m(\phi) := \left(\text{span}\{\phi_{m,n}(\cdot)\}_{n \in \mathbb{Z}} \right)^-$ é chamado de subespaço de aproximação, na escala m . Segundo Levan e Kubrusly (2003), a projeção ortogonal de $f(\cdot)$ sobre o subespaço de aproximação $V_m(\phi)$ é definida pela soma parcial, definida na equação (6).

$$f_{V_m(\phi)}(\cdot) := \sum_{n \in \mathbb{Z}} \langle f(\cdot), \phi_{m,n}(\cdot) \rangle \phi_{m,n}(\cdot). \quad (6)$$

De acordo com Daubechies (1988), a projeção $f_{V_m(\phi)}(\cdot)$ pode ser referida como uma componente de aproximação *wavelet* de $f(\cdot)$, na escala m , sobre $(V_m(\phi), \langle ; \rangle)$.

Uma transformada *wavelet* sobre l^2 é definida por um produto interno $\langle ; \rangle$ entre uma função $f(\cdot) \in l^2$ e uma função *wavelet* $\omega_{m,n}(\cdot)$ ou uma função escala $\phi_{m,n}(\cdot)$, onde $(m, n) \in \mathbb{Z} \times \mathbb{Z}$. De acordo com Morettin (2006), as transformadas *wavelet* podem ser agrupadas em dois conjuntos disjuntos: o de coeficientes de detalhe, denotado por $\{d_{m,n}\}_{(m,n) \in \mathbb{Z} \times \mathbb{Z}}$, e o de coeficientes de aproximação, denotado por $\{a_{m,n}\}_{(m,n) \in \mathbb{Z} \times \mathbb{Z}}$, onde um coeficiente de detalhe $d_{m,n}$ e um de aproximação $a_{m,n}$ são definidos, respectivamente, nas equações (7) e (8).

$$d_{m,n} := \langle f(\cdot), \omega_{m,n}(\cdot) \rangle = \sum_{t \in \mathbb{Z}} f(t) \omega_{m,n}(t), \quad (7)$$

$$a_{m,n} := \langle f(\cdot), \phi_{m,n}(\cdot) \rangle = \sum_{t \in \mathbb{Z}} f(t) \phi_{m,n}(t). \quad (8)$$

Baseado em Kubrusly e Levan (2006), a expansão de um mapa $f(\cdot)$ sobre o espaço de Hilbert l^2 , em termos de uma base ortonormal *wavelet*, é definida, genericamente, pela equação (9).

$$f(.) = \sum_{n \in \mathbb{Z}} a_{m_0,n} \phi_{m_0,n}(.) + \sum_{m=m_0}^{+\infty} \sum_{n \in \mathbb{Z}} d_{m,n} \omega_{m,n}(.), \quad (9)$$

onde $\sum_{n \in \mathbb{Z}} a_{m_0,n} \phi_{m_0,n}(.)$ é a componente *wavelet* de aproximação de nível m_0 e $\sum_{n \in \mathbb{Z}} d_{m,n} \omega_{m,n}(.)$ é a componente *wavelet* de detalhe de nível m do mapa $f(.)$.

2.2 ENCOLHIMENTO WAVELET

No caso da série temporal ser fortemente ruidosa, recomenda-se a filtragem da mesma, para então modelar a parte determinística obtida com a filtragem. Seja $(y_t)_{t=1}^T$ uma série temporal, onde para todo t : $y_t = f_t + e_t$. Sendo e_t um estado da variável aleatória $\varepsilon_t \sim (0, \sigma^2)$ e f_t a componente determinística de y_t . Como não é possível modelar e_t , indica-se o tratamento dos dados formadores da série temporal a fim de se obter a componente determinística. O encolhimento (*shrinkage*) *wavelet* é uma das técnicas destinada para este fim.

O Princípio do encolhimento *wavelet* tem como objetivo a redução (ou mesmo remoção) do ruído presente num sinal, diminuindo (ou zerando) a magnitude dos coeficientes *wavelet* (MORETTIN, 1999). Outras referências nesta área são os trabalhos de Donoho e Johnstone (1994) e Donoho *et al.*, (1995).

As três etapas do encolhimento *wavelet* são:

- 1) decomposição *wavelet* do sinal ruidoso;
- 2) selecionar limiares (*thresholds*) que eliminam (ou reduzam) os ruídos;
- 3) aplicar a transformada *wavelet* inversa a fim de se obter o sinal menos ruidoso.

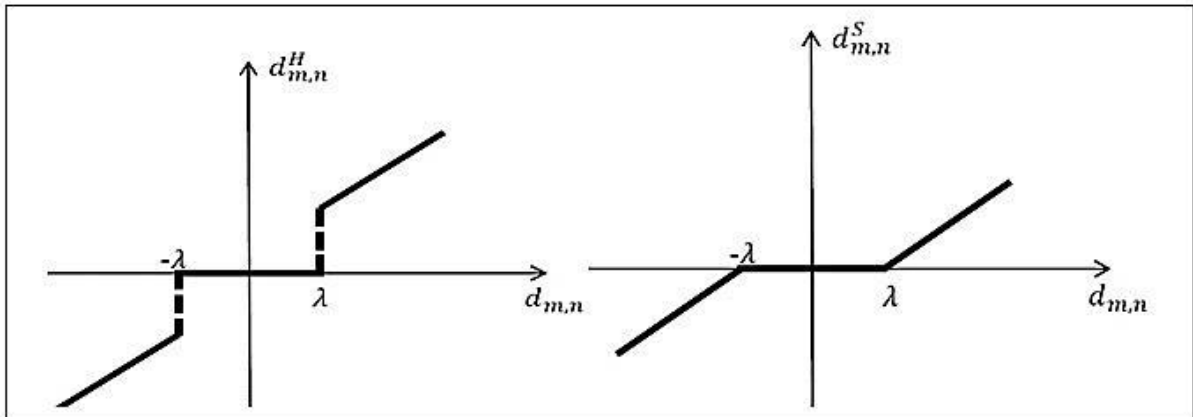
Das várias regras de limiarização encontradas na literatura, as mais tradicionais são o limiar duro (*hard threshold*) e limiar suave (*soft threshold*), representadas nas equações (10) e (11), respectivamente.

$$d_{m,n}^H = \begin{cases} d_{m,n}, & \text{se } |d_{m,n}| > \lambda \\ 0, & \text{se } |d_{m,n}| \leq \lambda \end{cases} \quad (10)$$

$$d_{m,n}^S = \begin{cases} \text{sign}(d_{m,n})(|d_{m,n}| - \lambda), & \text{se } |d_{m,n}| > \lambda \\ 0, & \text{se } |d_{m,n}| \leq \lambda \end{cases}, \quad (11)$$

onde $d_{m,n}$ é o coeficiente de detalhe com ruído, $d_{m,n}^H$ e $d_{m,n}^S$ são coeficientes de detalhes sem ruído, $\text{sign}(d_{m,n})$ é o sinal de $d_{m,n}$ e λ o limiar. O limiar duro gera uma descontinuidade, um “salto” em λ , eliminando ou mantendo o coeficiente de detalhe, enquanto que o limiar suave atenua por meio de $(|d_{m,n}| - \lambda)$ os valores de $d_{m,n}$ (se $|d_{m,n}| > \lambda$), anulando-os no caso de $|d_{m,n}| \leq \lambda$. Nos gráficos da Figura 3 podem ser observadas essas características.

FIGURA 3 - GRÁFICOS DAS FUNÇÕES LIMIAIR DURO E SUAWE.



FONTE: O autor (2015).

Neste trabalho o cálculo do limiar λ foi realizado a partir de duas propostas: universal e SURE (*Estimation Stein's Unbiased Risk Estimate*). Maiores detalhes dessas técnicas podem ser encontradas em Donoho e Johnstone (1993), Johnstone e Silverman (1996) e Morettin (1999).

De acordo com Donoho e Johnstone (1998) o limiar universal é calculado a partir da equação (12).

$$\lambda = \sigma \sqrt{2 \log(T)}, \quad (12)$$

sendo σ o nível do ruído e T a cardinalidade da série. O limiar universal é independente da escala m , e assume um valor global.

Donoho e Johnstone (1995) propuseram selecionar um limiar, pela minimização do estimador não-viesado do risco Stein (SURE), em cada escala m . Seja N_m o número de coeficientes na escala m , então o limiar de SURE é calculado a partir da equação (13), sendo o $SURE(d_m, \lambda)$ dado pela equação (14).

$$\lambda_m = \underset{0 \leq \lambda \leq \sigma_m \sqrt{2 \log(N_m)}}{\operatorname{argmin}} SURE(d_m, \lambda), \quad (13)$$

$$SURE(d_m, \lambda) = N_m - 2 \sum_{k=1}^{N_m} I(|d_{m,k}| \leq \lambda) + \sum_{k=1}^{N_m} \min(|d_{m,k}|, \lambda)^2, \quad (14)$$

onde σ_m é a medida de dispersão para cada nível m , $I(x) = \begin{cases} 1, & \text{se } x \text{ for verdadeiro} \\ 0, & \text{se } x \text{ for falso} \end{cases}$

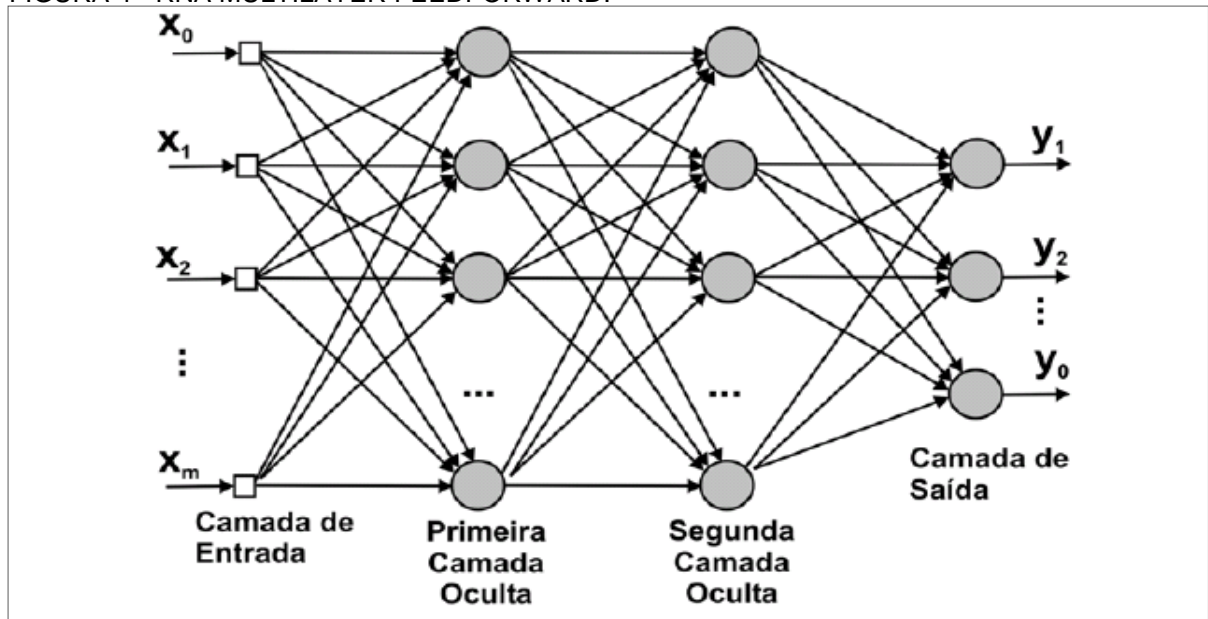
e $\hat{\sigma}_m = \text{mediana}(d_m)/0,6745$ um estimador de σ_m .

2.3 REDES NEURAIS ARTIFICIAIS

O mais eficiente sistema de aprendizado existente na natureza é o cérebro humano, constituído por neurônios, que são células capazes de armazenar e processar informações. Os neurônios estão ligados uns aos outros formando uma rede neural e o fluxo das informações (sinais) por esta rede depende de uma fenda existente entre os neurônios denominada de fenda sináptica, ou sinapse. Segundo Castro (2010), depois que todos os sinais de entrada são recebidos e ponderados pelas sinapses são somados no corpo celular, gerando um sinal de entrada líquida do neurônio e que será, posteriormente, propagado para a saída do neurônio.

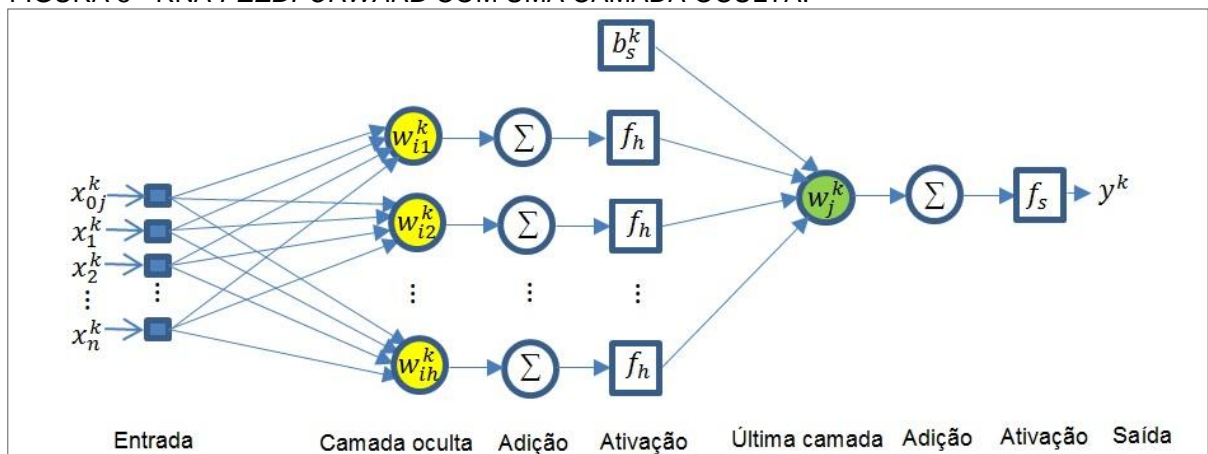
As *redes neurais artificiais* (RNA) foram idealizadas com base nos processos de aprendizagem do cérebro humano, de forma que uma RNA é capaz de simular as conexões sinápticas. De acordo com Haykin (2001), uma RNA é um sistema de processamento paralelo formado por unidades capazes de armazenar conhecimento e disponibilizá-lo para o futuro, sendo formada basicamente por neurônios artificiais distribuídos em camadas. As RNAs denominadas *multilayer feed-forward* são formadas por várias camadas e os sinais são propagados apenas da entrada para a saída, nunca ao contrário. A ligação entre os neurônios de uma camada e neurônios de camadas adjacentes ocorre por meio de pesos, que são parâmetros ajustáveis numericamente e que representam a “força” das conexões. O somatório das entradas linearmente ponderadas para um neurônio é transformado por uma função de ativação, que pode ser linear ou não linear (por exemplo, a tangente hiperbólica, a sigmoide). Neste tipo de RNA, não há conexões entre neurônios da mesma camada. As informações de dados são transmitidas de forma que as saídas dos neurônios de uma camada correspondam às entradas da camada seguinte, conforme ilustra a Figura 4.

FIGURA 4 - RNA MULTILAYER FEEDFORWARD.

FONTE: Barbosa *et al.*, (2005)

Na Figura 5, está representada uma RNA *feed-forward* com uma camada escondida, onde f_h e f_s denotam funções de ativação concernentes às junções aditivas oriundas da camada intermediária e da última camada, respectivamente. O sinal de entrada (padrão k) está definido pelo vetor de sinais $(x_1^k, x_2^k, \dots, x_n^k)$. Os pesos sinápticos das duas camadas foram denotados, respectivamente, por w_{ij}^k e w_j^k , com $i = 1, \dots, n$ e $j = 1, \dots, h$, sendo h o número de neurônios da camada oculta. As simbologias x_{0j}^k e b_s^k foram utilizadas para representar os bias, que podem aumentar ou diminuir a entrada líquida da função de ativação.

FIGURA 5 - RNA FEEDFORWARD COM UMA CAMADA OCULTA.



FONTE: O autor (2015).

Matematicamente, a saída y_k da rede representada na Figura 5 pode ser calculada através da equação (15).

$$y_k = f_s \left[\sum_{j=1}^h w_j^k f_h \left(\sum_{i=1}^n w_{ij}^k x_i^k + x_{0j}^k \right) + b_s^k \right]. \quad (15)$$

Para o treinamento da RNA, pode-se utilizar o algoritmo o *backpropagation*, ele é largamente utilizado e faz o ajuste dos pesos sinápticos por meio de um processo de otimização realizado em duas fases: *forward* e *backward*. Na fase *forward*, é calculada a resposta fornecida pela RNA para determinado padrão de entrada. Na fase *backward*, o desvio (ou erro) entre a resposta da RNA e a resposta desejada é utilizado no processo de ajuste dos pesos sinápticos. Ao longo do treinamento da RNA, os vários padrões de entrada e as respectivas respostas desejadas são apresentados à RNA, de forma que os pesos sinápticos sejam tais que minimizem a *média dos erros quadráticos* (MSE – *mean square error*) (TEIXEIRA Jr et al., 2012a).

Na sequência, são descritos, para a RNA ilustrada na Figura 5, os passos do algoritmo *backpropagation* padrão, que utiliza na otimização do erro o método do gradiente descendente. Neste algoritmo foi considerado o caso em que $f_h(\xi)$ e $f_s(\xi)$ são tais que $f_h(\xi) = f_s(\xi) = \frac{1}{1+e^{-\xi}}$ (função sigmoide). Inicialize os pesos das conexões e bias com valores aleatórios; defina a taxa de aprendizagem α , uma constante de proporcionalidade no intervalo [0,1]. Para cada padrão de entrada, execute os passos de I a III.

- (I) Calcule a saída da RNA: $z_j^k = f_h \left(\sum_{i=1}^n w_{ij}^k x_i^k + x_{0j}^k \right)$ e $y^k = f_s \left(\sum_{j=1}^h w_j^k z_j^k + b_s^k \right)$.
- (II) Faça correções de pesos e bias relativos à camada de saída: $w_j^k = w_j^k + \alpha y^k (1 - y^k) (d^k - y^k) z_j^k$ e $b_s^k = b_s^k + \alpha y^k (1 - y^k) (d^k - y^k)$, onde d^k é o valor desejado no padrão k .
- (III) Faça correções de pesos e bias relativos à camada escondida: $w_{ij}^k = w_{ij}^j + \alpha (d^k - y^k) y^k (1 - y^k) w_j^k z_j^k (1 - z_j^k) x_i^k$ e $x_{0j}^k = x_{0j}^k + \alpha (d^k - y^k) y^k (1 - y^k) w_j^k z_j^k (1 - z_j^k)$.

- (IV) Atualize a taxa de aprendizagem. Calcule os erros para todos os padrões de entrada, sendo o erro para determinado padrão dado por $E^k = \frac{1}{2}(d^k - y^k)^2$. Teste o critério de parada.

No campo das redes neurais artificiais, o algoritmo de treinamento de *Levenberg-Marquardt* (LM) é amplamente recomendado na resolução de problemas de pequeno e médio porte (HAYKIN, 2001). Este apresenta rápida convergência e estabilidade, ao contrário do *backpropagation* padrão que possui convergência lenta. O algoritmo LM é uma variação do algoritmo de Gauss-Newton, cuja atualização dos pesos é dada conforme a regra representada na equação (16)

$$w_{m+1} = w_m - (J'_m J_m)^{-1} J'_m e_m, \quad (16)$$

onde m é o índice da iteração e J_m a matriz jacobiana representada na equação (17).

$$J = \begin{bmatrix} \partial e_{11} / \partial w_1 & \partial e_{11} / \partial w_2 & \dots & \partial e_{11} / \partial w_p \\ \partial e_{12} / \partial w_1 & \partial e_{12} / \partial w_2 & \dots & \partial e_{12} / \partial w_p \\ \dots & \dots & \dots & \dots \\ \partial e_{1S} / \partial w_1 & \partial e_{1S} / \partial w_2 & \dots & \partial e_{1S} / \partial w_p \\ \dots & \dots & \dots & \dots \\ \partial e_{K1} / \partial w_1 & \partial e_{K1} / \partial w_2 & \dots & \partial e_{K1} / \partial w_p \\ \partial e_{K2} / \partial w_1 & \partial e_{K2} / \partial w_2 & \dots & \partial e_{K2} / \partial w_p \\ \dots & \dots & \dots & \dots \\ \partial e_{KS} / \partial w_1 & \partial e_{KS} / \partial w_2 & \dots & \partial e_{KS} / \partial w_p \end{bmatrix}. \quad (17)$$

O erro $e_{ks} = d_{ks} - y_{ks}$, $k = 1, \dots, K$, sendo K o total de padrões, $s = 1, \dots, S$, onde S é o total de saídas da rede. Os valores desejados e saídas atuais da rede estão representados por d_{ks} e y_{ks} , respectivamente. A fim de garantir a invertibilidade da matriz hessiana $(J'_m J_m)$, o algoritmo LM introduz o termo μI , onde μ é uma constante positiva e I a matriz identidade. De forma que a atualização dos pesos no algoritmo LM ocorre conforme Equação (18).

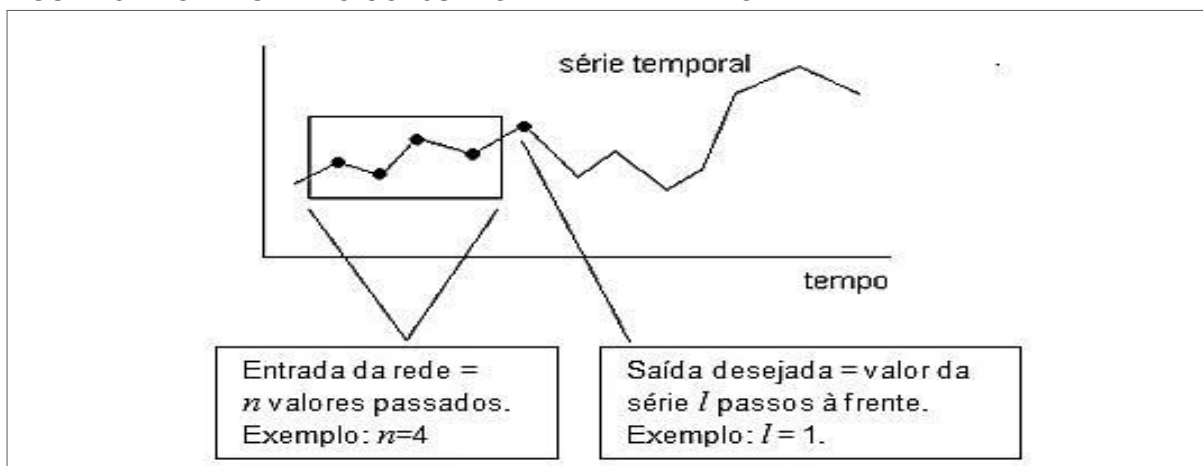
$$w_{m+1} = w_m - (J'_m J_m + \mu I)^{-1} J'_m e_m. \quad (18)$$

O problema de previsão de séries temporais, com T observações $(y_t)_{t=1}^T$, através de RNA consiste na utilização de dados observados para se prever $y(t + l)$,

onde l é o passo de tempo de previsão para o futuro. Uma vez definido o tamanho da janela de entrada de dados (n) e o valor de l , costroi-se a sequência $(y(t), y(t-1), \dots, (y(t-n+1), y(t+1), \dots, y(t+l)))$ de dados, que forma um conjunto de padrões de treinamento, onde: $(y(t), y(t-1), \dots, y(t-n+1))$ e $(y(t+1), \dots, y(t+l))$ representam as entradas e saídas desejadas, respectivamente.

A Figura 6 ilustra um caso particular, onde $n = 4$ e $l = 1$. A janela é deslocada ao longo da série até que se tenha $T - 4$ padrões de entrada/saída.

FIGURA 6 - MONTAGEM DO CONJUNTO DE TREINAMENTO.



FONTE: Teixeira Jr. *et al.*, (2012a) – Adaptado.

2.4 COMBINAÇÃO DE PREVISÕES

Tomadores de decisão, por vezes, possuem mais de uma previsão para uma mesma variável e a escolha de uma delas pode refletir de julgamentos subjetivos que são na maioria dos casos imprecisos. Questiona-se então, qual é a melhor estratégia de escolha de previsão entre várias? A combinação, linear ou não linear, de previsões individuais é largamente utilizada, pois geram resultados positivos. Embora haja um consenso entre os pesquisadores de que a combinação de previsores individuais, em regra, incorre em ganhos preditivos na projeção de séries temporais, não existe um senso comum em se definir a melhor forma de se fazê-la (WALLIS, 2011).

De acordo com Clemen (1989), a combinação de previsões consiste em uma abordagem clássica que utiliza algum mecanismo ponderador (calculado de acordo

com algum objetivo) associado às previsões a serem combinadas. Geralmente, adota-se, como objetivo, a minimização da estatística MSE (*mean square error*).

Segundo Granger (1989), a combinação de previsões, de fato, tende a gerar previsões mais acuradas. De acordo com Hollauer *et al* (2008), a diversificação de previsões leva naturalmente à diminuição dos erros de previsão. Faria e Mubwandarikwa (2008) salientam que a previsão combinada pode ser interpretada como uma agregadora de informações oriundas de diferentes fontes (no caso, os métodos preditivos individuais) sobre a flutuação estocástica da série temporal a ser projetada. Em Wallis (2011), é feita uma revisão histórica, na qual são mencionados muitos dos principais artigos sobre combinação de métodos preditivos individuais, publicados em quarenta e dois anos, os quais atestam a sua utilização e a sua eficiência, além de mostrarem uma enorme diversidade de aplicações e abordagens.

Os precursores na área das combinações de previsões são Bates e Granger (1969), que sugeriram uma combinação, expressa na equação (19), para a previsão da série temporal $(y_t)_{t=1}^T$.

$$\hat{y}_t = \rho_1 \times \hat{y}_{1,t} + \rho_2 \times \hat{y}_{2,t}. \quad (19)$$

Na equação (19), a previsão linearmente combinada é representada por \hat{y}_t , enquanto as previsões obtidas através de dois métodos, denotados por 1 e 2 são $\hat{y}_{1,t}$ e $\hat{y}_{2,t}$, respectivamente. Além disso, é assumido que $\rho_1 + \rho_2 = 1$ (para preservar a interpretabilidade probabilística), sendo que os seus valores ótimos são calculados a partir da minimização dos erros das variâncias das previsões combinadas, equação (20)

$$\sigma_c^2 = \rho_1^2 \sigma_1^2 + \rho_2^2 \sigma_2^2 + 2\rho\rho_1\rho_2\sigma_1\sigma_2, \quad (20)$$

onde ρ é o coeficiente de correlação dos erros das previsões, σ_1 e σ_2 as variâncias dos erros das previsões individuais e ρ_1 dado pela equação (21).

$$\rho_1 = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \quad (21)$$

Gupta e Wilton (1987) usaram em seus trabalhos a média aritmética das previsões individuais, os resultados de uma experiência de simulação mostraram

que o método é altamente eficiente, e significativamente superior em relação a outras abordagens existentes, em muitas condições. De forma que a equação (22) fornece a previsão combinada.

$$\hat{y}_t = \frac{\rho_1 \times \hat{y}_{1,t} + \rho_2 \times \hat{y}_{2,t} + \dots + \rho_n \times \hat{y}_{n,t}}{n}. \quad (22)$$

Granger e Ramanathan (1984) propuseram, entre outros métodos, uma combinação linear sem restrição para os coeficientes e adição de uma constante. Terui e Dijk (2002) sugerem o uso de filtros de Kalman para o cálculo dos coeficientes e que estes sejam variáveis ao longo do tempo. Concluíram que resultados das previsões combinadas são superiores às previsões provenientes de métodos individuais. Por outro lado, mais pesquisas são necessárias para se verificar a robustez dos resultados.

Teixeira Jr *et al.*, (2012b) resolveu um problema de programação matemática para calcular os coeficientes da combinação linear. Obtiveram ganhos preditivos ainda maiores na amostra de treino, aumentando o poder de generalização das previsões fora da amostra.

2.4.1 Otimização

São duas as abordagens clássicas para problemas de otimização: escalar e vetorial. De acordo com Izmailov e Solodov (2009a) um problema geral de otimização, na abordagem escalar, pode ser representado na forma expressa em (23).

$$\text{minimizar } f(x) \text{ sujeito a } x \in D. \quad (23)$$

Este problema pode ser classificado em linear ou não linear dependendo das funções usadas na sua construção. Sendo D o conjunto viável do problema e $f(x)$ a função objetivo. O problema de minimização formulado em (23) pode ser transformado em um problema de maximização desde que se troque o sinal da função objetivo, ou seja: *maximizar* $-f(x)$ *sujeito a* $x \in D$. O conjunto viável D , por sua vez, é determinado por funções ditas de restrição, então (23) pode ser reformulado e escrito na forma (24).

$$\begin{aligned} & \text{minimizar } f(\mathbf{x}) \\ \text{sujeito a: } & \begin{cases} g_i(\mathbf{x}) \leq 0, & i = 1, \dots, p \\ h_i(\mathbf{x}) = 0, & i = 1, \dots, q \end{cases} \end{aligned} \quad (24)$$

Considere (24) um problema de otimização não linear, de forma que pelo menos uma das funções que constituem o problema seja não linear. Em um problema de otimização não linear irrestrito com função objetivo côncava tem-se a garantia de que um máximo local é um máximo global. De forma similar, se a função objetivo é convexa tem-se a garantia que o mínimo local é mínimo global. Em um problema de otimização não linear restrito, como (24), o máximo (mínimo) global é garantido se a função objetivo é côncava (convexa) e a região de soluções viáveis é um conjunto convexo.

No caso de (24) possuir apenas as restrições de igualdade, o problema pode ser resolvido a partir dos Multiplicadores de Lagrange. Considere o problema de maximizar ou minimizar a função $f(x_1, x_2, \dots, x_n)$ sujeita às condições $h_1(x_1, x_2, \dots, x_n) = c_1, h_2(x_1, x_2, \dots, x_n) = c_2 \dots h_q(x_1, x_2, \dots, x_n) = c_q$, com $q \leq n - 1$. Se $(x_1^*, x_2^*, \dots, x_n^*)$ é solução do problema e se $f(x_1, x_2, \dots, x_n)$ e $h_i(x_1, x_2, \dots, x_n)$, $i = 1, \dots, q$ possuem derivadas parciais de primeira ordem contínuas em uma vizinhança de $(x_1^*, x_2^*, \dots, x_n^*)$, então existem multiplicadores $\lambda_1^*, \lambda_2^*, \dots, \lambda_q^*$ tais que $(x_1^*, x_2^*, \dots, x_n^*, \lambda_1^*, \lambda_2^*, \dots, \lambda_q^*)$ deve ser um ponto crítico da função de Lagrange $F(x_1, x_2, \dots, x_n, \lambda_1, \lambda_2, \dots, \lambda_q) = f(x_1, x_2, \dots, x_n) - \sum_{i=1}^q \lambda_i (h_i(x_1, \dots, x_n) - c_i)$. O ponto crítico $\mathbf{x}^* = (x_1^*, x_2^*, \dots, x_n^*, \lambda_1^*, \lambda_2^*, \dots, \lambda_q^*)$ deve satisfazer a seguinte igualdade: $\frac{\partial F}{\partial x_1} = \frac{\partial F}{\partial x_2} = \dots = \frac{\partial F}{\partial x_n} = \frac{\partial F}{\partial \lambda_1} = \frac{\partial F}{\partial \lambda_2} = \dots = \frac{\partial F}{\partial \lambda_q} = 0$.

A existência de restrições de desigualdades em (24) implica no uso de outros otimizadores além dos Multiplicadores de Lagrange. Entre tantas as opções existentes na literatura, segue breve descrição do método das penalidades e barreiras.

- (I) **Método das penalidades:** reescreve-se o problema (24) na formulação $\text{minimizar } F(\mathbf{x}, \boldsymbol{\mu}) = f(\mathbf{x}) + Q(\mathbf{x})$, sendo $Q(\mathbf{x})$ uma função de penalidades que é formada por $g(\mathbf{x})$ e $h(\mathbf{x})$ e escrita da seguinte maneira $Q(\mathbf{x}) = r_h \left(\sum_{j=1}^q (h_j(\mathbf{x}))^2 \right) + r_g \left(\sum_{i=1}^p (\max\{0, g_i(\mathbf{x})\})^2 \right)$, onde $r \geq 0$ e é atualizado pelo produto $r_{k+1} = r_k C$, com $C > 1$. Assim, r tende para

infinito no processo iterativo, sendo k a ordem da iteração. Violação na restrição de igualdade implica em penalidade do termo $r_h[h_i(x)]^2$. Se $g(x) \leq 0$, a expressão $\max\{0, g_i(x)\}$ é nula, caso contrário tem-se a penalidade $r_g[g(x)]^2$. O processo inicia-se em um ponto x_0 que pode estar na região viável ou não.

- (II) **Método das barreiras:** Considere no problema (24) a ausência das restrições de igualdade. Assim, ele pode ser escrito na forma *minimizar* $F(x, \mu) = f(x) + \sum_{i=1}^p \frac{-\mu_i}{g_i(x)}, \mu > 0$. Onde μ deve ser decrescente e o seu valor em iteração k é dado pelo produto de C e μ da iteração $k - 1$, sendo C um valor do intervalo $(0,1)$. O algoritmo inicia-se com a escolha de um ponto x (solução inicial) qualquer que pertença à região viável do problema inicial e termina quando $\sum_{i=1}^p \frac{-\mu_i}{g_i(x)} < \varepsilon$ (tolerância).

Um método de otimização bastante utilizado é o Gradiente Reduzido Generalizado (GRG). Neste método as variáveis são separadas em básicas e não básicas, em cada iteração o método considera apenas as variáveis não básicas, sendo o gradiente em relação a estas variáveis chamado de gradiente reduzido. Para achar a direção viável o método faz a projeção do gradiente da função objetivo nas restrições ativas. Uma variante deste método é o GRG2 utilizado pelo *Solver* do Excel na resolução de problemas não lineares (COLIN, 2007).

Funções de otimização do Matlab (*fmincon*, *fminimax*, etc) utilizam o algoritmo Programação Quadrática Sequencial (PQS). Neste método, obtêm-se equações não lineares aplicando as condições de Karush-Kuhn-Tucker no Lagrangiano do problema restrito. As equações não lineares obtidas são resolvidas pelo método de Newton.

Maiores detalhes sobre otimização podem ser encontrados em Izmailov e Solodov (2009a; 2009b).

2.5 BOOTSTRAP

O *bootstrap* é um procedimento estatístico não paramétrico introduzido por Efron (1979). A técnica constitui na reamostragem de mesmo tamanho e com reposição dos dados da amostra original, com o objetivo de se estimar uma medida

de interesse. O princípio básico do método *bootstrap* estabelece que na ausência de informações sobre a verdadeira distribuição de uma variável aleatória, pode-se obter uma distribuição empírica a partir de reamostragens de uma amostra observada. Considerando-se que esta amostra possui informações da distribuição base, B amostras daquela permitirão estimar parâmetros da população da qual foi retirada.

Considere $X = (X_1, X_2, \dots, X_n)$ uma amostra aleatória de uma distribuição F e $\hat{\theta} = s(X)$ estimador de um parâmetro θ desta distribuição. Uma amostra *bootstrap* $X^* = (X_1^*, X_2^*, \dots, X_n^*)$ é obtida da amostra original X rearranjando os seus n elementos, podendo ocorrer valores repetidos. Assim, uma das possibilidades para X^* é $X^* = (X_1^* = X_2, X_2^* = X_n, \dots, X_n^* = X_1)$, fixando-se a probabilidade de escolha em X de $1/n$ para cada uma das suas observações. A partir deste procedimento determinam-se B amostras *bootstrap* independentes da distribuição empírica \hat{F} , obtendo-se então o estimador *bootstrap* $\hat{\theta}^{*(b)} = s(X^{*(b)})$, com $b = 1, 2, \dots, B$, do parâmetro θ . O erro padrão do estimador *bootstrap* pode ser representado através da equação (25).

$$\widehat{se}_{boot}(\hat{\theta}) = \left[\frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}^{*(b)} - \hat{\theta}^*)^2 \right]^{1/2}, \quad (25)$$

onde $\hat{\theta}^* = \frac{1}{B} \sum_{b=1}^B \hat{\theta}^{*(b)}$.

De acordo com Montgomery e Runger (2003), a definição do número B de amostras *bootstrap* pode ser realizada com a observação da variação do desvio padrão do estimador *bootstrap*. A baixa variabilidade ou estabilidade desse valor indica o valor de B mais adequado.

Com relação à previsão de séries temporais, segundo Souza e Camargo (2004), existem duas maneiras para se fazer *bootstrap* em uma série de tempo: (a) *bootstrap* nos resíduos; (b) método *moving blocks*. O primeiro é mais usado que o segundo para aplicações de *bootstrap* em séries temporais.

Mais detalhes sobre a técnica *bootstrap* podem ser encontrados em Freedman e Peters (1984), Efron e Tibshirani (1986), Souza e Neto (1996), dentre outros.

2.6 INTERVALO DE CONFIANÇA E PREVISÃO

A aplicação de modelos de regressão implica na avaliação das incertezas que cercam estimadores de parâmetros e variáveis aleatórias. Os intervalos de confiança e previsão são duas medidas estatísticas valiosas que desempenham esta importante função. O primeiro mede as variações concernentes aos parâmetros do modelo de regressão, enquanto o segundo é mais amplo e considera não só as variações relativas ao modelo de regressão, mas também as variações consequentes dos erros preditivos.

Considere os padrões $\{(y_t, \mathbf{x}_t)\}_{t=1}^T$ e o modelo de regressão neural $y(\mathbf{x}) = f(\mathbf{x}, \mathbf{w}) + e(\mathbf{x})$, onde $y(\mathbf{x})$ são os alvos, \mathbf{w} são os parâmetros verdadeiros do modelo desconhecido $f(\mathbf{x}, \mathbf{w})$ e $e(\mathbf{x})$ o ruído (variável aleatória) com média zero. Denotando por $g(\mathbf{x}, \hat{\mathbf{w}})$, onde $\hat{\mathbf{w}}$ são os estimadores de \mathbf{w} , a aproximação de $f(\mathbf{x}, \mathbf{w})$, sendo esta aproximação interpretada como a média dos alvos (y) dada a entrada \mathbf{x} .

De acordo com Heskes (1997), para problemas de regressão podem-se distinguir dois diferentes aspectos: a acurácia da estimativa da regressão verdadeira e a acurácia das estimativas relativas às saídas observadas. O primeiro aspecto conduz à construção do intervalo de confiança, resultado da distribuição da quantidade $f(\mathbf{x}, \mathbf{w}) - g(\mathbf{x}, \hat{\mathbf{w}})$, o segundo aspecto trata da variação $y(\mathbf{x}) - g(\mathbf{x}, \hat{\mathbf{w}})$ e leva ao intervalo de previsão, conforme equação (26).

$$y(\mathbf{x}) - g(\mathbf{x}, \hat{\mathbf{w}}) = f(\mathbf{x}, \mathbf{w}) - g(\mathbf{x}, \hat{\mathbf{w}}) + e(\mathbf{x}). \quad (26)$$

Da equação (26) tem-se que o intervalo de confiança está contido no intervalo de previsão. Segundo Khosravi *et al.* (2014) a variância total associada ao resultado da modelagem pode ser expresso como na equação (27).

$$\sigma_y^2(\mathbf{x}) = \sigma_g^2(\mathbf{x}) + \sigma_e^2(\mathbf{x}), \quad (27)$$

onde $\sigma_g^2(\mathbf{x})$ é a incerteza do modelo de regressão ajustado e $\sigma_e^2(\mathbf{x})$ a medida da variância do ruído $e(\mathbf{x})$.

Tibshirani (1996) sugere, no contexto das regressões, dois tipos de amostragens *bootstrap* que permitem determinar o intervalo de confiança: *bootstrap pairs* e *bootstrap residual*.

A seguir, um resumo dos passos que constituem o algoritmo da amostragem *bootstrap pairs*:

- (1) Gerar B amostras ($20 \leq B \leq 200$) cada uma com tamanho n , obtidas com reamostragem *bootstrap* das n observações de treinamento $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$. Denotando as B amostras *bootstrap* por $\{(x_1^{*b}, y_1^{*b}), (x_2^{*b}, y_2^{*b}), \dots, (x_n^{*b}, y_n^{*b})\}$, com $b = 1, \dots, B$.
- (2) Para cada amostra *bootstrap*, minimize $\sum_{i=1}^n [y_i^{*b} - g(x_i^*, w)]^2$ para obter \hat{w}^{*b} .
- (3) Estimar o erro padrão do i -ésimo valor predito usando $\left\{ \frac{1}{B-1} \sum_{b=1}^B [g(x_i, \hat{w}^{*b}) - m(x_i)]^2 \right\}^{1/2}$ onde $m(x_i) = \frac{1}{B} \sum_{b=1}^B g(x_i, \hat{w}^{*b})$.

O algoritmo do *bootstrap residual* apresenta alguns passos em comum com o *bootstrap pairs*. A principal diferença reside na construção das séries *bootstrapadas*. O procedimento para execução do *bootstrap residual* está descrito a seguir.

- (1) Utilizar os dados observados para determinar as estimativas dos parâmetros do modelo neural \hat{w} , fazendo $r_i = y_i - g(x_i, \hat{w})$, $i = 1, \dots, n$.
- (2) Gerar B amostras *bootstrap*, cada uma com tamanho n e obtidas com a reamostragem de r_1, r_2, \dots, r_n . Sejam $r_1^{*b}, r_2^{*b}, \dots, r_n^{*b}$ as B amostras *bootstrap*, com $b = 1, 2, \dots, B$. Faça $y_i^{*b} = g(x_i, \hat{w}) + r_i^{*b}$.
- (3) Para cada amostra *bootstrap*, minimize $\sum_{i=1}^n [y_i^{*b} - g(x_i^*, w)]^2$, determinando \hat{w}^{*b} .
- (4) Estimar o erro padrão do i -ésimo valor predito usando $\left\{ \frac{1}{B-1} \sum_{b=1}^B [g(x_i, \hat{w}^{*b}) - m(x_i)]^2 \right\}^{1/2}$, onde $m(x_i) = \frac{1}{B} \sum_{b=1}^B g(x_i, \hat{w}^{*b})$.

Khosravi *et al.*, (2014), trabalharam com amostragem *bootstrap* e RNA. Segundo esses autores o **intervalo de predição** pode ser expresso pela inequação (28).

$$m(\mathbf{x}) - t_{1-\frac{\alpha}{2}, gl} \sqrt{\sigma_y^2(\mathbf{x})} \leq y(\mathbf{x}) \leq m(\mathbf{x}) + t_{1-\frac{\alpha}{2}, gl} \sqrt{\sigma_y^2(\mathbf{x})}, \quad (28)$$

onde gl são graus de liberdade e α o nível de significância em uma distribuição t de Student. Os autores definiram gl como sendo a diferença entre o número de amostras de treinamento e o número de parâmetros do modelo RNA. Seguindo nesta linha, o **intervalo de confiança** pode ser escrito na forma representada na inequação (29).

$$m(\mathbf{x}) - t_{1-\frac{\alpha}{2}, gl} \sqrt{\sigma_g^2(\mathbf{x})} \leq f(\mathbf{x}, \mathbf{w}) \leq m(\mathbf{x}) + t_{1-\frac{\alpha}{2}, gl} \sqrt{\sigma_g^2(\mathbf{x})}. \quad (29)$$

Carney; Cunningham; Bhagwan (1999) obtiveram intervalo de confiança e previsão a partir de um *ensemble* (combinação de diversos resultados de previsão) de RNA com *bootstrap*. Os autores sugerem dividir o conjunto com B redes *bootstrap* em M subconjuntos de tamanho β , executando os seguintes passos:

(1) Determinar $m^{*i}(\mathbf{x})$ em cada um dos M subconjuntos formados por β redes e montar o conjunto $E = \{m^{*i}(\mathbf{x})\}_{i=1}^M$.

(2) Obter B^* reamostragens *bootstrap* de E , formando o conjunto $G = \{E_j^*\}_{j=1}^{B^*}$, onde $E_j^* = \{m^{*j1}(\mathbf{x}), \dots, m^{*jM}(\mathbf{x})\}$.

(3) Calcular a variância em cada um dos B^* conjuntos E_j^* :

$$\sigma_j^{2*} = \frac{1}{M} \sum_{k=1}^M (m^{*jk}(\mathbf{x}) - m^{*j}(\mathbf{x}))^2.$$

(4) Fazer a média das variâncias: $\sigma_g^2 = \frac{1}{B^*} \sum_{j=1}^{B^*} \sigma_j^{2*}$.

Com uma melhor estimativa para variância da distribuição $P(f(\mathbf{x}, \mathbf{w})|m^*(\mathbf{x}))$, constrói-se o IC para $f(\mathbf{x}, \mathbf{w})$, conforme a inequação (30).

$$m^*(\mathbf{x}) - z_{1-\alpha} \sigma_g \leq f(\mathbf{x}, \mathbf{w}) \leq m^*(\mathbf{x}) + z_{1-\alpha} \sigma_g, \quad (30)$$

onde $z_{1-\alpha}$ provêm da distribuição normal padrão com nível de significância α e $m^*(\mathbf{x})$ é obtido a partir das B amostragens *bootstrap*.

Nix; Weigend (1995) sugeriram uma técnica para estimar a média e a variância de uma distribuição de probabilidades de alvos em um problema de regressão por redes neurais artificiais. Os autores utilizaram para este fim a máxima verossimilhança para formar a função custo, descrita na equação (31), para treinar uma rede neural.

$$C = \sum_i \frac{1}{2} \left[\frac{(d_i - y(x_i))^2}{\sigma_e^2(x_i)} + \ln(\sigma_e^2(x_i)) \right], \quad (31)$$

onde $d(x)$ são os valores observados (alvos) e $y(x) \approx f(x)$ no modelo $d(x) = f(x) + e(x)$. Esta proposta é frequentemente adaptada para calcular a variância do ruído ($\sigma_e^2(x)$), sendo utilizada pelos autores Heskes (1997), Carney; Cunningham; Bhagwan (1999), Khosravi *et al.* (2014) dentre outros.

Bartkiewicz (2012) realizou previsões através de redes neuro-fuzzy e estimou a variância do ruído ($\sigma_e^2(x)$) treinando uma rede com padrões $\{x_i, |e_i|\}, i = 1, \dots, n$, sendo $|e_i| = |y_i - g(x_i, \hat{w})|$. O autor comparou o intervalo de previsão determinado através da sua proposta com intervalos oriundos de outras técnicas, e concluiu que os resultados foram satisfatórios.

2.7 META-HEURÍSTICA PSO

A otimização por enxame de partículas PSO – *Particle Swarm Optimization* foi criada por James Kennedy e Russel Eberhart em 1995 após estudos da convivência social dos pássaros. De acordo com os pesquisadores os pássaros trocam informações na busca de alimento, onde cada membro de um bando de pássaros determina sua velocidade com base em sua experiência pessoal, bem como informações obtidas por meio da interação com outros membros do bando. Cada pássaro, denominado partícula, de início voa pelo espaço de busca e a partir da cooperação social desloca-se na direção do alimento.

Matematicamente, cada partícula é considerada uma possível solução para um problema de otimização e os vetores \mathbf{X} e \mathbf{V} representam a sua posição e velocidade, respectivamente. A atualização de velocidade e posição de cada partícula é realizada através das seguintes fórmulas.

$$V_{ij}^{k+1} = WV_{ij}^k + c_1 r_{1ij} (p_{ij}^k - X_{ij}^k) + c_2 r_{2ij} (g_j^k - X_{ij}^k), i = 1 \dots, m \text{ e } j = 1, \dots, n \quad (32)$$

$$X_{ij}^{k+1} = X_{ij}^k + V_{ij}^{k+1}, i = 1, \dots, m, \quad (33)$$

onde: m é o número de partículas, n representa a dimensão do problema, c_1 e c_2 representam os coeficientes cognitivos individuais e sociais, respectivamente, r_{1ij} e r_{2ij} números aleatórios entre 0 e 1 e k denota a iteração. Da equação (32) o termo WV_{ij}^k simboliza a inércia, e conduz a partícula para a direção que vinha seguindo. O segundo termo $c_1 r_{1ij} (p_{ij}^k - X_{ij}^k)$ simboliza a memória, atraindo a partícula para o melhor ponto encontrado durante a sua trajetória. O terceiro termo $c_2 r_{2ij} (g_j^k - X_{ij}^k)$ simboliza a cooperação, que conduz as partículas para o melhor ponto até então encontrado pelo enxame.

O algoritmo básico do PSO, segundo Behnamian e Ghomi (2010), segue os passos:

Passo 1: Inicializar uma população de partículas com posições aleatórias e velocidades, onde cada partícula contém n variáveis.

Passo 2: Avaliar a função objetivo para todas as partículas. Fazer p_{ij}^k igual à melhor posição de cada partícula. Fazer g_j^k e seu valor objetivo ser igual à posição e ao valor objetivo da melhor partícula inicial.

Passo 3: Atualizar a velocidade e a posição de cada partícula de acordo com as equações (32) e (33).

Passo 4: Avaliar os valores objetivo de todas as partículas.

Passo 5: Para cada partícula, comparar o seu valor objetivo atual com o valor objetivo da sua melhor posição. Se o valor atual é melhor, atualize p_{ij}^k e seu valor objetivo com a posição e valor objetivo atual.

Passo 6: Determinar a melhor partícula do enxame atual com o melhor valor objetivo. Se o valor objetivo é melhor do que o valor objetivo de g_j^k atualize g_j^k e seu valor objetivo com a posição e o valor objetivo da melhor partícula corrente.

Passo 7: Se o critério de parada (número de iterações) for atendido, informe g_j^k e seu valor objetivo, caso contrário, vá para ao passo 3.

O PSO descrito deve sofrer adaptações quando usado na otimização de problemas discretos. A primeira sugestão de adaptação foi apresentada por Kennedy e Eberhart (1997). Neste método as partículas são representadas por

sequências binárias e a velocidade em termos de probabilidades. A forma para calcular a velocidade permanece inalterada e a posição dada pela equação (34).

$$X_{ij}^{k+1} = \begin{cases} 1, & \text{se } \text{sig}(V_{ij}^{k+1}) > r_{ij}, \\ 0, & \text{caso contrário} \end{cases}, \quad (34)$$

onde, $\text{sig}(V_{ij}^{k+1}) = \frac{1}{1+\exp(-V_{ij}^{k+1})}$. Na equação (34) r_{ij} é um número randômico do intervalo $[0,1]$.

Behnamian e Ghomi (2010) propuseram um método híbrido envolvendo *particle swarm optimization* (PSO) e *simulated annealing* (SA) para estimar os parâmetros de um modelo de regressão não linear, compararam os resultados obtidos com os métodos ARIMA e SARIMA. Concluíram que o método proposto produziu erros inferiores.

Teixeira, Siqueira e Teixeira Jr (2014) utilizaram PSO e Algoritmo Genético (AG) no ajuste de modelos ARMA. O Algoritmo Genético foi utilizado na determinação das ordens p (autorregressão) e q (médias móveis). A meta-heurística PSO, por sua vez, foi usada na obtenção dos parâmetros autorregressivos e médias móveis. Os resultados obtidos mostram que o modelo ajustado é adequado à série estudada.

Yu, Wang e Xi (2008) apresentaram um método de previsão por RNA, com utilização de PSO na otimização conjunta da estrutura de uma RNA *feedforward* e parâmetros (pesos e bias). Experimentos com dados reais mostraram a utilização da meta-heurística PSO levou à RNAs compactas e com boa capacidade de generalização.

3 MATERIAIS E MÉTODOS

Para esta tese foram realizados dois experimentos computacionais envolvendo previsões de séries temporais. No primeiro deles (experimento 1), fizeram-se previsões pontuais das séries *Canadian Lynx*, *Wolf's Sunspot* e *Exchange Rate*. No segundo experimento foram estimados os intervalos de confiança e previsão para uma série de vazões mensais. Neste capítulo são apresentados os materiais utilizados e os métodos empregados nos dois experimentos.

3.1 MATERIAS

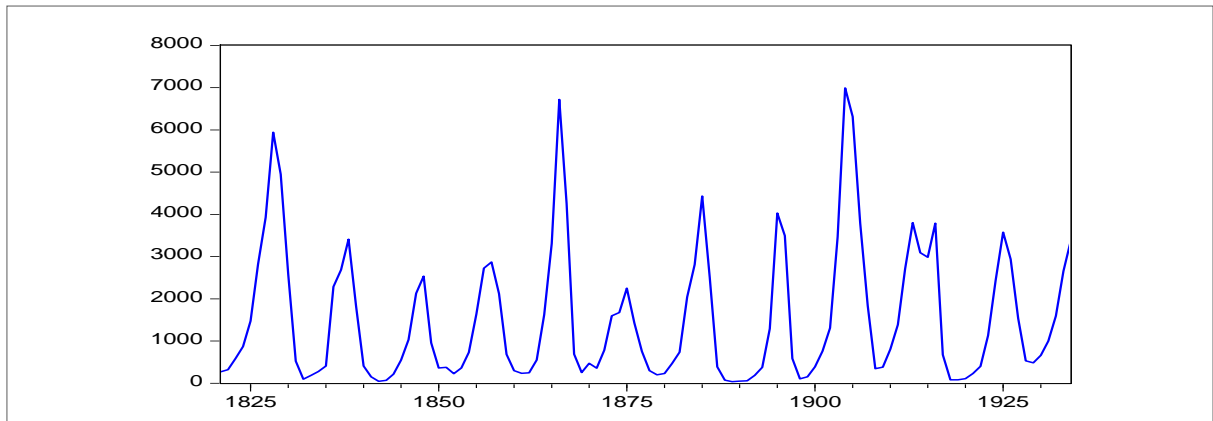
Nesta seção encontram-se informações relativas às séries temporais utilizadas nos experimentos 1 e 2. As séries temporais do experimento 1 podem ser obtidas em <https://datamarket.com/>. A série de vazões mensais (experimento 2), pode ser visualizada em http://www.ons.org.br/operacao/vazoes_naturais.aspx.

3.1.1 Série temporal *Canadian Lynx* - Experimento Computacional 1

A série de dados *Canadian Lynx* refere-se ao registro anual do número de *Lynx canadensis* (espécie de felino natural da América do Norte) capturados no distrito de Rio Mackenzie no Norte do Canadá. Existem na literatura vários trabalhos que fazem análise e previsão da série de tempo *Canadian Lynx*, credita-se a Moran (1953) o primeiro trabalho de análise desta série. Outras abordagens aplicadas na análise e previsão da série dos lincos canadenses podem ser encontradas nos trabalhos de: Zhang (2003); Subba e Sabr (1984); Tang & Ghosal (2007); Wong e Li (2000), entre outros.

O gráfico da série *Canadian Lynx* (114 observações no período: 1821-1934) está representado na Figura 7, onde se pode observar uma periodicidade de aproximadamente 10 anos (STONE e HE, 2007).

FIGURA 7 - GRÁFICO DA SÉRIE TEMPORAL CANADIAN LYNX NO PERÍODO 1821-1934

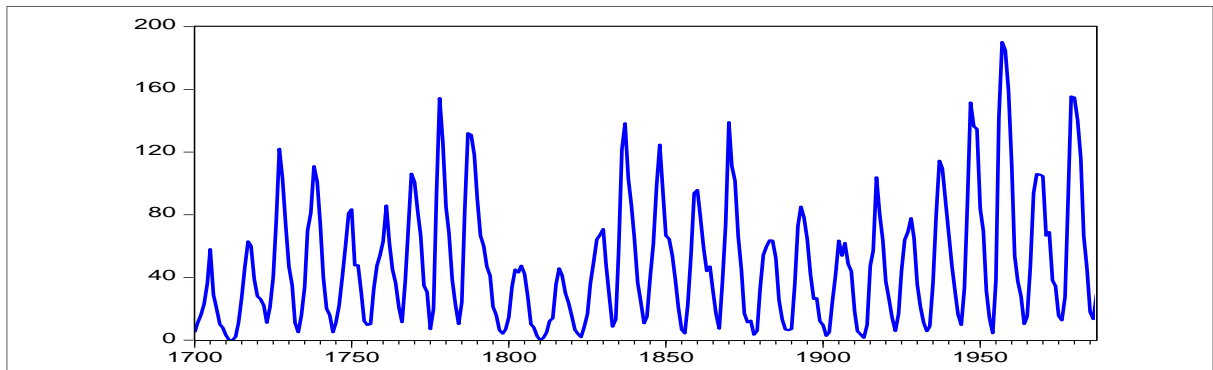


FONTE: O autor (2015).

3.1.2 Série Temporal *Wolf's Sunspot* – Experimento Computacional 1

As manchas solares são geradas por fortes campos magnéticos criados no interior do sol e são de interesse de climatologistas e geofísicos (ECHER, *et al.* 2003; ZHANG, 2003). Em 1854, Rudolf Wolf desenvolveu um método para quantificar o número de manchas solares e construiu uma série de médias anuais (*Wolf's Sunspot*), com início no ano de 1700. Esta série temporal de dados é considerada não linear e não-Gaussiana e muitas vezes utilizada para avaliar a eficácia de modelos não lineares (ZHANG, 2003). A série temporal *Wolf's Sunspot* apresenta um padrão cíclico com ciclo médio de cerca de onze anos. Na Figura 8, é representado gráfico da *Wolf's Sunspot*, com 288 observações, para os anos de 1700 até 1987.

FIGURA 8 - GRÁFICO DA SÉRIE TEMPORAL WOLF'S SUNSPOT NO PERÍODO 1700-1987

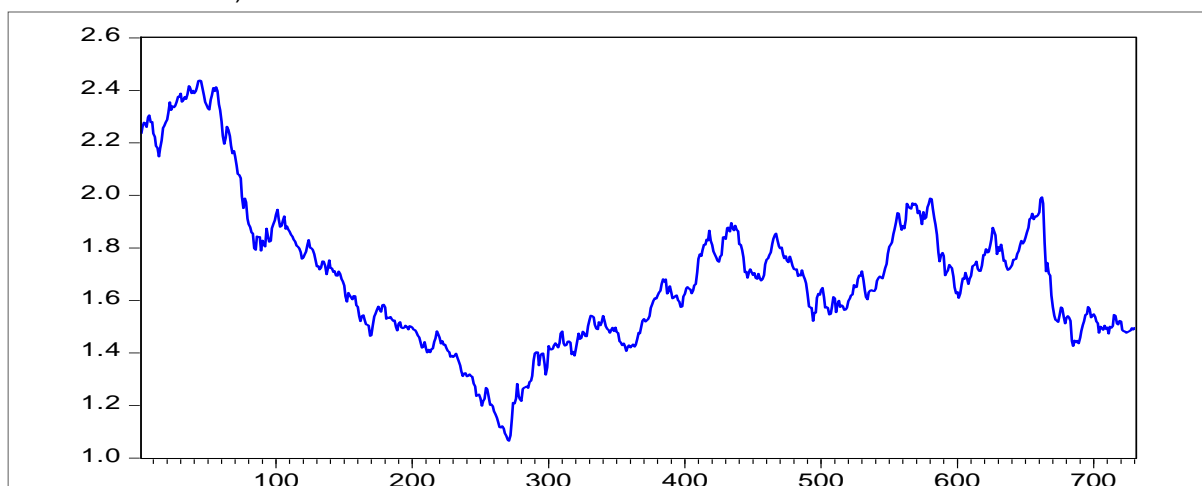


FONTE: O autor (2015).

3.1.3 Série temporal *Exchange Rate* – Experimento Computacional 1

Uma taxa de câmbio é o preço de uma moeda em termos de outra moeda. Neste trabalho, a taxa de câmbio envolve as moedas libra esterlina e dólar americano. No mercado financeiro internacional, fazer a previsão da taxa de câmbio é bastante importante, mas de difícil execução. Neste sentido, vários modelos lineares e não lineares foram desenvolvidos para a realização de tal tarefa, com algumas aplicações apresentando resultados contraditórios (KHASHEI; BIJARI, 2011). A série temporal *Exchange Rate* utilizada na modelagem possui 731 pontos e abriga observações semanais de 1980 a 1993. O gráfico da série pode ser observado na Figura 9.

FIGURA 9 - GRÁFICO DA SÉRIE TEMPORAL *EXCHANGE RATE* COM 731 OBSERVAÇÕES (1980-1993)



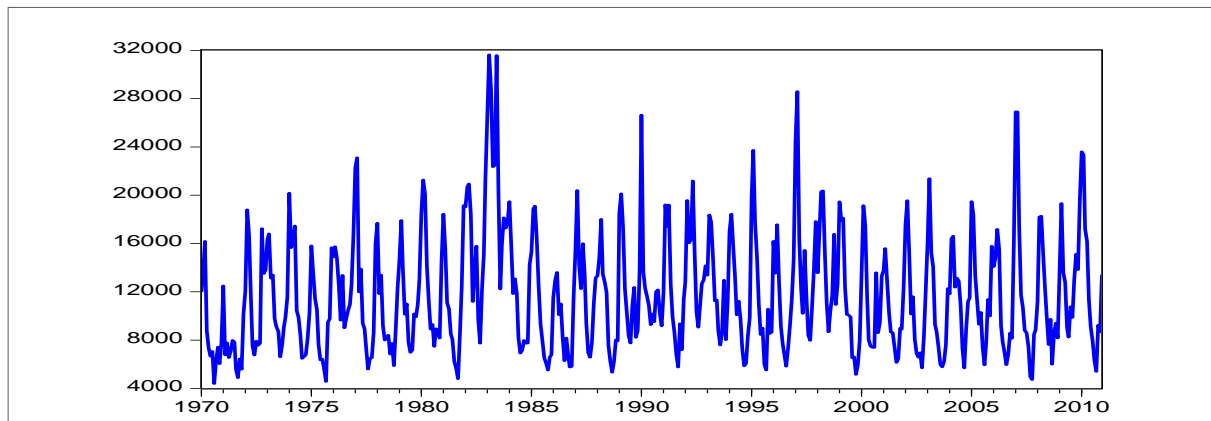
FONTE: O autor (2015).

3.1.4 Série de Vazões Mensais – Experimento Computacional 2

No experimento computacional 2, foi utilizada uma série temporal de vazão de afluentes mensal média da usina de Itaipu cujo período é de janeiro de 1970 até dezembro de 2010 (ou seja, 492 observações). Mais especificamente, os dados são provenientes das medições no posto 266, da Usina Hidrelétrica de Itaipu, localizado no Rio Paraná, no trecho de fronteira entre o Brasil e o Paraguai, a 14 km ao norte da Ponte da Amizade. A área da usina se estende desde Foz do Iguaçu, no Brasil, e Ciudad del Este, no Paraguai, ao sul, até Guaíra (Brasil) e Salto del Guairá (Paraguai), ao norte.

A escolha dessa série se deu por conta de sua relevância e por ser de difícil modelagem. Na Figura 10, tem-se o gráfico da série temporal supracitada, com todas as 492 observações.

FIGURA 10 - SÉRIE DE VAZÕES MENSAIS OBTIDAS NO POSTO 266-ITAIPU NO PERÍODO 1970-2010



FONTE: O autor (2015).

3.2 MÉTODOS

Nesta seção são apresentados os métodos empregados nas previsões das séries temporais dos experimentos 1 e 2. Os métodos individuais ARIMA e RNA e os compostos *Wavelet*-ARIMA e *Wavelet*-RNA foram utilizados na previsão da série de vazões (experimento 2). O método proposto (capítulo 4) foi utilizado em ambos os experimentos.

3.2.1 Experimento Computacional 1

A fim de avaliar a qualidade do método proposto em previsões pontuais (seções 4.1, 4.2 e 4.3), procedeu-se a sua aplicação nas previsões das séries de tempo *Canadian Lynx*, *Wolf's Sunspot* e *Exchange Rate* e os seus resultados estatísticos comparados com os alcançados por Zhang (2003); Kajitani, Keith e Mcleod (2005); Aladag, Egrioglu e Kadilar, (2009); Adhikari e Agrawal, (2012); Khashei e Bijari, (2010, 2011 e 2012) e Ismail e Shabri (2014).

As três séries temporais foram divididas em amostras de treino, validação e teste. Os tamanhos destas amostras e as descrições das três séries de tempo utilizadas neste experimento e pelos autores citados no parágrafo anterior foram resumidas e representadas na Tabela 1.

TABELA 1 - DESCRIÇÃO DAS SÉRIES *CANADIAN LYNX*, *WOLF'S SUNSPOT* E *EXCHANGE RATE* E TAMANHO DAS AMOSTRAS UTILIZADAS NA MODELAGEM.

Séries Temporais	Descrição	Amostras		
		Treino	Validação	Teste
<i>Canadian Lynx</i>	Registro anual do número de lincas capturados no distrito do Rio Mackenzie no norte do Canadá (1821-1934).	80	20	14
<i>Wolf's Sunspot</i>	Número de manchas solares observadas anualmente (1700-1987).	173	48	67
<i>Exchange Rate</i>	Taxa de câmbio semanal entre libra inglesa e dólar americano (1980-1993).	512	168	52

FONTE: Adhikari e Agrawal (2012) adaptado.

NOTA: as séries originais *Canadian Lynx* e *Exchange Rate* foram transformadas por logaritmo decimal.

Como estatísticas de aderência, foram utilizados o MSE (*mean square error*) e o MAE (*mean absolute error*), definidos, respectivamente, abaixo, em (35).

$$MSE = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_t)^2 \quad e \quad MAE = \frac{1}{T} \sum_{t=1}^T |y_t - \hat{y}_t|. \quad (35)$$

É comum entre os autores que publicaram trabalhos de previsão da série dos lincas canadenses e taxa de câmbio (utilizada neste trabalho) aplicarem logaritmo de base dez ao conjunto de dados originais. Fazendo-se, para as três séries, previsões um passo à frente. Os horizontes de previsões para as séries *Canadian Lynx*, *Wolf's Sunspot* e *Exchange Rate* são 14 anos, 67 anos e 52 semanas, respectivamente.

3.2.2 ARIMA

Segundo Box e Jenkins (1970) um modelo ARMA plausível para a série temporal $\{y_t\}_{t=1}^T$, com cardinalidade T , de ordem p (autorregressivo - AR) e q (médias móveis - MA) é descrito pela equação (36).

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} + a_t \quad (36)$$

O modelo (36) combina valores passados das entradas y_t e choques aleatórios (a_t) descorrelacionados, de média zero e variância constante. Neste modelo $\phi_i, \theta_j \in \mathbb{R}$, com $i = 1, \dots, p$ e $j = 1, \dots, q$, denotam os parâmetros do modelo e δ uma constante. Na hipótese da série temporal ser não estacionária, esta deve ser diferenciada e o modelo ARMA(p, q) substituído pelo ARIMA(p, d, q), sendo d a

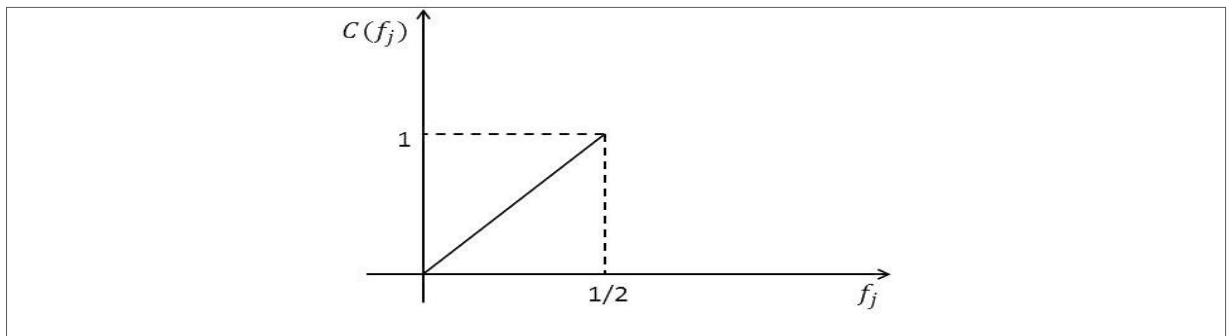
ordem de diferenciação da série. Para a identificação do modelo, as ordens p e q podem ser determinadas por meio da análise do perfil dos gráficos das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) (HAMILTON 1994). Identificado o modelo, passa-se ao estágio seguinte que é a estimação dos parâmetros. Para tanto, é necessário utilizar métodos iterativos não lineares de mínimos quadrados. Maiores detalhes destas aplicações podem ser encontrados em Box e Jenkins (1970) e Morettin e Tolo (2006). Para a validação do modelo já com os parâmetros estimados pode-se usar testes estatísticos, tais como: teste de Box-Pierce, teste do periodograma acumulado, teste da autocorrelação residual, entre outros. No caso da série temporal $\{y_t\}_{t=1}^T$ apresentar componente sazonal, o modelo de Box & Jenkins plausível é o ARIMA multiplicativo, dado genericamente na equação (37).

$$\phi(B)(1 - \phi_1 B - \dots - \phi_p B^p) \nabla^d (1 - B^S)^D y_t = \theta(B)(1 - \theta_1 B - \dots - \theta_q B^q) a_t \quad (37)$$

onde: $\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$, $\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$, d é a ordem das diferenças simples, D é a ordem das diferenças sazonais, S é o período sazonal, $\phi_k \in \mathbb{R}$ e $\theta_j \in \mathbb{R}$ são os coeficientes dos polinômios não sazonais e $\phi_m \in \mathbb{R}$ e $\theta_n \in \mathbb{R}$ são os coeficientes dos polinômios sazonais.

A fim de validar o modelo ajustado foram analisados os resíduos com a construção dos gráficos das funções FAC e FACP, buscando mostrar que os mesmos são não autocorrelacionados. A validação é reforçada com o periodograma integrado, segundo Morettin e Tolo (2006) se o modelo não for adequado, haverá desvios sistemáticos da reta no gráfico do periodograma acumulado $C(f_j)$ pela frequência (f_j) , veja Figura 11.

FIGURA 11 - PERIODOGRAMA ACUMULADO



FONTE: O autor (2015).

Para testar a hipótese que todos os coeficientes de autocorrelação ρ_k são iguais a zero, foi utilizada a estatística $Q^* = n \sum_{k=1}^m \hat{\rho}_k^2$, desenvolvida por Box e Pierce, em que n é o tamanho da amostra e m a duração da defasagem. A estatística Q^* tem distribuição qui-quadrado com m graus de liberdade. Quando Q^* excede o valor crítico, rejeita-se a hipótese nula de que todos os ρ_k são iguais a zero (GUJARATI, 2000).

3.2.3 Redes Neurais Artificiais *Feedforward*

A RNA foi empregada na modelagem das séries temporais dos experimentos computacionais 1 e 2, nas formas: (a) individual; (b) associada com a decomposição *wavelet*; e (c) no método proposto (capítulo 4). Nas três abordagens a RNA utilizada foi do tipo *feedforward multilayer perceptron* (RNA-MLP), constituída por uma camada escondida com função de ativação tangente hiperbólica sigmoidal (*tansig*) ou linear e um neurônio na camada de saída com função de linear (*purelin*) ou exponencial. Para os treinamentos das RNAs aplicou-se o algoritmo de *Levenberg & Marquardt* e a meta-heurística PSO.

Na abordagem individual, utilizada na previsão das vazões médias mensais, Foram realizadas vinte combinações envolvendo tamanho da janela de entrada (J), com J=4,5,6 e 7 e quantidade de neurônios na camada escondida (CE), sendo CE=4,5,6,7 e 8, fazendo-se dez simulações para cada combinação.

A série das vazões foi dividida de tal forma que, reservou-se 70% para o treinamento, 20% para validação e 10% para teste. A melhor configuração da RNA foi escolhida mediante o menor Erro Médio Percentual Absoluto (MAPE), equação (38), na amostra de validação.

$$MAPE = \frac{1}{T} \sum_{t=1}^T \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad (38)$$

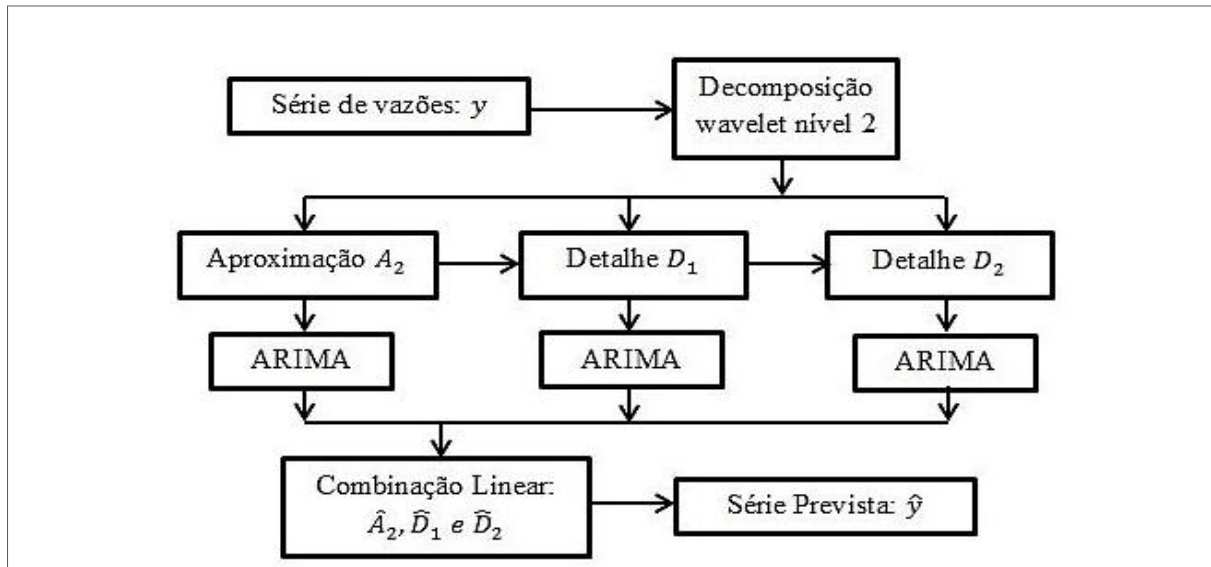
3.2.4 Wavelet-ARIMA e Wavelet-RNA

Os métodos compostos formados a partir da associação da decomposição *wavelet* com ARIMA e RNA foram aplicados na previsão da série temporal de

vazões mensais (experimento 2). Inicialmente, a série de vazões foi decomposta ortogonalmente via *wavelet* de nível $p=2$. As três componentes (duas de detalhes e uma de aproximação) obtidas com a decomposição foram ajustadas por três RNA e três modelos ARIMA e as previsões combinadas linearmente a fim de se obter a série observada. Utilizou-se para as previsões por ARIMA o aplicativo *EViews* 8 e o Matlab 8.0 para a modelagem por RNA, com dez simulações para cada combinação entre janela ($J=10, \dots, 22$) e número de neurônios na camada escondida ($CE=5, \dots, 10$).

Na Figura 12 podem ser observadas as etapas para a execução do método *Wavelet*-ARIMA. O fluxograma do método *Wavelet*-RNA pode ser obtido pela substituição, na Figura 12, do ARIMA pela RNA.

FIGURA 12 - FLUXOGRAMA COM AS ETAPAS DO MÉTODO *WAVELET*_ARIMA



FONTE: O autor (2015).

Sejam $\hat{A}_{2,RNA}$, $\hat{D}_{1,RNA}$ e $\hat{D}_{2,RNA}$ as previsões das componentes *wavelet* de aproximação e detalhes 1 e 2, respectivamente, obtidas por RNA. Estas componentes são combinadas linearmente a fim de se obter a previsão da série de vazões, conforme equação (39).

$$(\hat{y}_t)_{t=1}^T = \alpha_{A2,RNA} \hat{A}_{2,RNA} + \alpha_{D1,RNA} \hat{D}_{1,RNA} + \alpha_{D2,RNA} \hat{D}_{2,RNA}, \quad (39)$$

onde $\alpha_{A2,RNA}$, $\alpha_{D1,RNA}$ e $\alpha_{D2,RNA}$ são parâmetros a serem determinados com a resolução do problema de programação não linear (40), cuja função objetivo é a raiz do erro quadrático médio (RMSE).

$$\text{Min } RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (\hat{y}(t) - y(t))^2} \quad (40)$$

s. a. $\alpha_{A2,RNA}, \alpha_{D1,RNA}$ e $\alpha_{D2,RNA}$ são irrestritas.

De forma análoga obtém-se $(\hat{y}_t)_{t=1}^T$ a partir das previsões oriundas do modelo ARIMA: $\hat{A}_{2,ARIMA}, \hat{D}_{1,ARIMA}$ e $\hat{D}_{2,ARIMA}$.

4 MÉTODO PROPOSTO

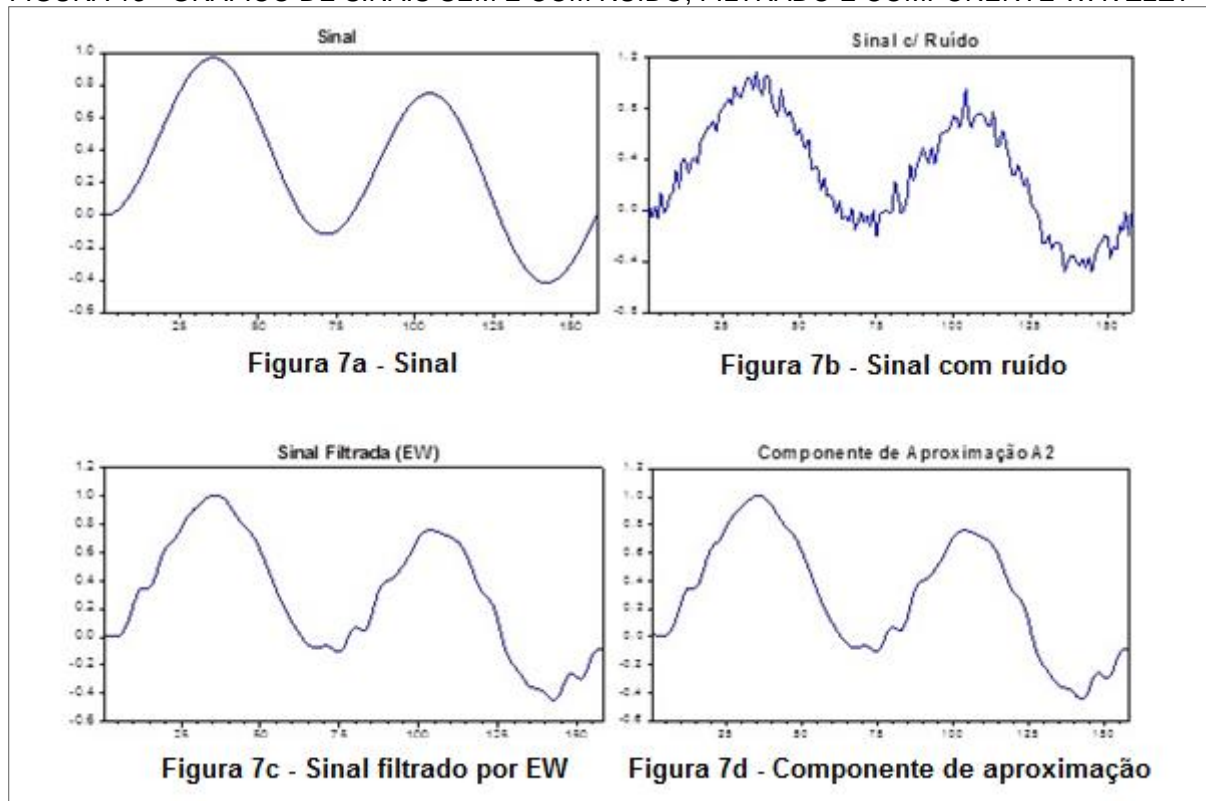
Neste capítulo são descritas as etapas do método proposto, sendo elas: submissão da série observada ao encolhimento *wavelet*, decomposição *wavelet* das séries obtidas com o encolhimento, modelagem das séries temporais X e Z provenientes da decomposição *wavelet* via RNA, geração de cenários por *bootstrap*, combinação linear dos cenários e determinação dos intervalos de confiança e previsão.

4.1 ENCOLHIMENTO WAVELET DA SÉRIE OBSERVADA

A Transformada *Wavelet* Discreta (TWD) é definida para sinais com comprimento igual a potência de dois, uma consequência da decomposição diádica. Sinais que não possuem esta característica podem ser completados por simetrização, preenchimento com zeros e suave, entre outros (PROTÁZIO, 2002). De acordo com Ariño (1995), os algoritmos usados na TWD e reconstrução do sinal levam a efeitos de borda, sendo os coeficientes *wavelet* de fronteira diferentes em relação aos teóricos. O mesmo autor propõe, após a DW, excluir os primeiros e últimos termos das componentes *wavelet*. Desta forma, antes da filtragem *wavelet*, adiciona-se a série temporal $\{y_t\}_{t=1}^T$, onde T totaliza as cardinalidades das amostras de treino e validação, q informações de forma que $T + q = 2^Q$, $Q \in \mathbb{Z}_+^*$. Ariño (1995) sugere que as q informações sejam previsões da série $\{y_t\}_{t=1}^T$ q passos à frente e obtidas por ARIMA, RNA, etc. No caso das séries periódicas, além das previsões, sugere-se completar $\{y_t\}_{t=1}^T$ com as médias em cada instante dentro do período, no caso do tamanho do período ser inferior a q adiciona-se as últimas informações da série y_t . Obtendo-se a série $\{y_t\}_{t=1}^{T+q}$, que é submetida ao encolhimento *wavelet*, dando origem a uma série temporal de tendência $\{y_{t,EW}\}_{t=1}^{T+q}$ e uma componente total de resíduo $\{y_{t,R}\}_{t=1}^{T+q} = \{y_t\}_{t=1}^{T+q} - \{y_{t,EW}\}_{t=1}^{T+q}$. A componente de tendência $\{y_{t,EW}\}_{t=1}^{T+q}$ resultante da filtragem *wavelet* é proveniente da eliminação e/ou atenuação dos coeficientes de detalhes (seção 2.2). A Figura 13 ilustra a situação onde os coeficientes de detalhes são todos (ou quase todos) eliminados. O sinal com ruído (Figura 13b) foi submetido ao encolhimento *wavelet* (EW) com decomposição pela base daubechies de ordem 10 (db 10) nível 2, limiar universal e regra de limiarização

suave, resultando o sinal filtrado (Figura 13c). Na Figura 13d está representado o gráfico da componente de aproximação resultante da decomposição db10 nível 2 do sinal com ruído. Comparando as Figuras 13c e 13d, vê-se que parte dos coeficientes de detalhes foram eliminados, dada a semelhança entre as duas figuras.

FIGURA 13 - GRÁFICO DE SINAIS SEM E COM RUÍDO, FILTRADO E COMPONENTE *WAVELET*



FONTE: O autor (2015).

Resultados empíricos obtidos com o emprego das bases daubechies, symlet e coiflet no encolhimento *wavelet*, mostraram que as bases com menores níveis de decomposição e ordens, eram mais indicadas na busca pela acurácia das previsões das séries estudadas. Desta forma, a busca pela base *wavelet* a ser usada no EW ficou restrita às ordens $N=1,2,\dots,10$ e níveis de decomposição $p = 1,\dots,4$.

À componente de resíduo $y_{t,R} = y_t - y_{t,EW}$ são aplicados os testes de autocorrelação residual, independência estatística BDS e de estacionariedade *Dickey-Fuller*. Com o teste BDS (Brock *et al.*, 1987) pode-se investigar se os resíduos são independentes e identicamente distribuídos (i.i.d), o teste tem como hipótese nula a classificação do resíduo como i.i.d e hipótese alternativa a presença de dependência linear ou não linear. O teste BDS utiliza o conceito de correlação

especial dos termos da série dentro de um espaço de dimensão “m”. De acordo com Lima (2011), o teste BDS deve ser executado da segunda até a sexta dimensão, portanto, a série de resíduos rejeitará a hipótese nula se, e somente se, pelo menos um dos cinco níveis descritos no teste (p-valor) for menor do que o nível de significância adotado no teste com uma distribuição normal padrão. A verificação de estacionariedade de uma série temporal pode ser feita a partir do teste da raiz unitária, onde a hipótese nula diz que a variável estocástica possui raiz unitária (série temporal não estacionária), enquanto a hipótese alternativa é não ter raiz unitária (série temporal estacionária). Dickey Fuller é um dos testes de raiz unitária e a sua aplicação é válida quando o processo gerador da série temporal é um processo autorregressivo de ordem 1 (AR(1)) com termo de erro aleatório classificado como ruído branco (LIMA, 2011). Considerando o modelo AR(1): $y_t = \phi_1 y_{t-1} + a_t$, onde a_t são os erros aleatórios, se $\phi_1 = 1$, diz-se que a variável estacionária y tem raiz unitária. Uma série temporal que tenha raiz unitária é não estacionária (GUJARATI, 2000). No caso de séries temporais mais complexas que um AR(1), aplica-se o teste do Dickey Fuller aumentado (LIMA, 2011).

A submissão da componente de resíduo $y_{t,R}$ aos testes estatísticos mencionados no parágrafo anterior, tem como finalidade obter-se indicação que a mesma não pode ser classificada estatisticamente como ruído branco, de forma que os resíduos devem possuir estruturas de autodependência linear e/ou não linear que podem ser modeladas. Neste caso, a previsão da série observada resultará da soma das previsões dos resíduos e tendência ($\hat{y}_t = \hat{y}_{t,EW} + \hat{y}_{t,R}$), observando que, para a previsão um passo a frente \hat{y}_{T+1} as q informações são descartadas, utilizando no treinamento e validação a série temporal de tamanho T .

4.2 DECOMPOSIÇÃO WAVELET: TENDÊNCIA E RESÍDUO

Considere a sequência infinita $\{\tilde{y}_{t,S}\}_{t \in \mathbb{Z}}$, onde $S = EW$ ou R , obtida da série temporal $\{y_{t,S}\}_{t=1}^T$ a partir da igualdade $\tilde{y}_{t,S} = \begin{cases} y_{t,S}, & t \in \{1, \dots, T\} \\ 0, & t \in \mathbb{Z} \setminus \{1, \dots, T\} \end{cases}$, sendo $\tilde{y}_{t,S}$ quadraticamente somável e pertencente ao espaço de Hilbert $(l^2, \langle \cdot, \cdot \rangle)$. Segundo Kubrusly e Levan (2006); Mallat (1998), l^2 admite a expansão ortogonal representada na equação (41).

$$l^2 = V_{m_0}(\phi) + \left(\sum_{m=m_0}^{\infty} W_m(\omega) \right)^-, \quad (41)$$

onde $V_{m_0}(\phi)$ é um subespaço fechado de aproximação de nível m_0 e $W_m(\omega)$ um subespaço fechado de detalhe de nível m . Segue que a decomposição *wavelet* da série temporal $\tilde{y}(\cdot) \in l^2$ pode ser escrita segundo a expansão representada na equação (42).

$$\tilde{y}(\cdot) = \sum_{n \in \mathbb{Z}} a_{m_0,n} \phi_{m_0,n} + \sum_{m=m_0}^{\infty} \left(\sum_{n \in \mathbb{Z}} d_{m,n} \omega_{m,n} \right), \quad (42)$$

onde $\{\phi_{m_0,n}\}_{n \in \mathbb{Z}} \cup \{\omega_{m,n}\}_{m \geq m_0}$ é uma base ortonormal *wavelet* de do espaço l^2 , $d_{m,n} = \langle \tilde{y}(\cdot), \omega_{m,n} \rangle$ e $a_{m_0,n} = \langle \tilde{y}(\cdot), \phi_{m_0,n} \rangle$.

Segundo Teixeira Jr *et al.* (2012b) a decomposição *wavelet* de nível p de uma série temporal $\tilde{y}_{t,S}$ em relação a uma determinada base ortonormal *wavelet* é representada na equação (43).

$$\tilde{y}_{t,S} \cong \tilde{\tilde{y}}_{t,S} = \sum_{n \in \Gamma_{V_{m_0}}} a_{m_0,n} \phi_{m_0,n} + \sum_{m=m_0}^{m_0+(p-1)} \left(\sum_{n \in \Gamma_{W_m}} d_{m,n} \omega_{m,n} \right), \quad (43)$$

onde $\Gamma_{V_{m_0}}$ e Γ_{W_m} são subconjuntos de índices finitos inclusos em \mathbb{Z} . A determinação de tais conjuntos pode ocorrer através de diversas abordagens, geralmente baseadas na minimização do erro quadrático médio (MSE) entre $\tilde{y}_{t,S}$ e $\tilde{\tilde{y}}_{t,S}$.

Representando as componentes de aproximação e detalhe por $\tilde{y}_{V_{m_0}(\phi)}$ e $\tilde{y}_{W_m(\omega)}$, respectivamente, a equação (43) pode ser reeditada na forma da equação (44).

$$\tilde{y}_{t,S} \cong \tilde{y}_{V_{m_0}(\phi)} + \sum_{m=m_0}^{m_0+(p-1)} \tilde{y}_{W_m(\omega)}. \quad (44)$$

As componentes de aproximação e detalhe são obtidas a partir do aplicativo computacional Matlab 8.0, que oferece várias opções de bases *wavelets*, tais como:

as famílias daubechies (db N), coiflets (coif N) e symlets (sym N) (N representa a ordem da decomposição) desenvolvidas por Ingrid Daubechies (DAUBECHIES, 1992). O Quadro 1 traz algumas características dessas famílias.

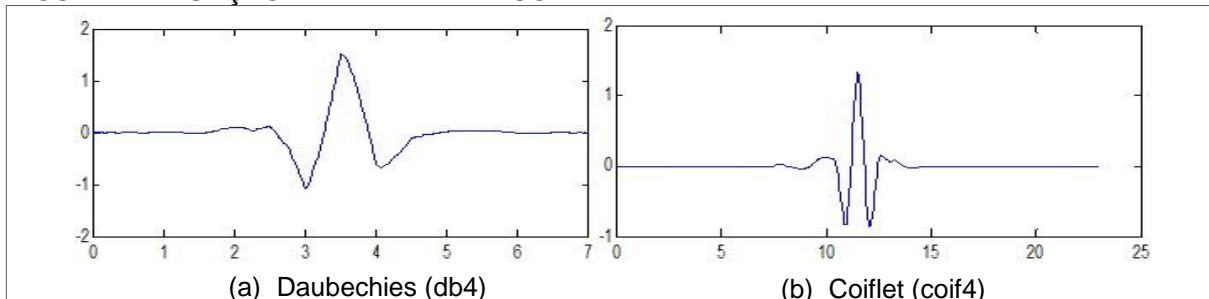
QUADRO 1 - ALGUMAS CARACTERÍSTICAS DAS FAMÍLIAS DB, SYM E COIF.

Características	Famílias <i>Wavelet</i>		
	db N	sym N	coif N
Ordem	N=1,2,...	N=1,2,..	N=1,2,...,5
Ortogonal	Sim	Sim	Sim
Suporte compacto	Sim	Sim	Sim
Largura do suporte	2N-1	2N-1	6N-1
Simetria	Aproxima-se	Aproxima-se	Aproxima-se
Número de momentos nulos	N	N	2N

FONTE: O autor (2015).

A representação gráfica das funções *wavelet* daubechies e coiflets para a ordem N=4 estão expostas na Figura 14.

FIGURA 14 - FUNÇÃO WAVELET DB4 E COIF4



FONTE: O autor (2015).

Não há na literatura um consenso com relação a melhor regra para escolha da base *wavelet* e, muitos autores não mencionam em seus trabalhos o procedimento adotado na escolha da base, levando a intuir que as escolhas são empíricas. Em Torrence e Compo (1998) encontra-se uma discussão do tema. Esses autores fazem análises das características de bases *wavelet* e séries temporais que podem ajudar em uma escolha qualitativa da base mais apropriada. Rivas, Burgos e Garcia-Prada (2009) baseiam-se no teorema de Parseval, que compara a energia do sinal à energia presente nos coeficientes *wavelet*, para apresentar o critério $1/T \sum [|f(t)|]^2 = \sum_{-\infty}^j \sum [|d_j|]^2 + \sum |a_j|^2$, sendo d_j e a_j os coeficientes wavelet de detalhes e aproximação, respectivamente, e $f(t)$ o sinal.

As bases *wavelet* da família daubechies são largamente aplicadas no estudo de séries temporais, pois apresentam alta capacidade de análise e síntese. Devido à regularidade destas bases, aproximam eficazmente funções suaves do espaço l^2 . Então, para a série de vazões mensais, escolheu-se para a decomposição ortogonal *wavelet* das séries de tendência e resíduos EW as bases de Daubechies db N em nível p , com o valor de N definido a partir do erro quadrático médio entre as séries decompostas e as respectivas somas das $p + 1$ componentes *wavelet* (TEIXEIRA, 2013). As séries de tempo do experimento computacional 1 também foram decomposta a partir das bases daubechies, com exceção da *Wolf's Sunspot*, onde resultados empíricos levaram a escolha da base symlet.

Considerando-se a série temporal $\{y_{t,EW}\}_{t=1}^{T+q}$ obtida na seção 4.1, calcula-se a sua média (μ) e obtém-se a série temporal centrada em zero por meio da diferença: $\{Y_{t,EW}\}_{t=1}^{T+q} = \{y_{t,EW}\}_{t=1}^{T+q} - \mu$. Em seguida, faz-se decomposição *wavelet* (DW) de nível p de $\{Y_{t,EW}\}_{t=1}^{T+q}$. Os coeficientes $d_{m,n}$, nos primeiros níveis de resolução representam a escala fina (detalhe dos dados), enquanto os coeficientes $d_{m,n}$ nos últimos níveis de resolução, juntamente com os coeficientes $a_{m_0,n}$ representam a escala mais grossa (parte suave dos dados) (MORETTIN, 1999). Assim, podem-se determinar séries temporais X e Z a partir dos coeficientes da escala grossa e fina, respectivamente. Ariño (1995) decompôs via *wavelet*, em nível $p = 7$, uma série econômica e agrupou as componentes *wavelet* obtidas em dois conjuntos de dados, onde cada um deles refletia a série observada em diferentes frequências.

A fim de se determinar os níveis de resolução que comporão X e Z , calcula-se a energia $E(m) = \sum_{n=0}^{2^m-1} d_{m,n}^2$, $m = 0, 1, \dots, Q - 1$, dos coeficientes wavelet de detalhes ($d_{m,n}$). Sejam π e Π níveis de resolução das componentes de altas e baixas frequências, respectivamente, onde ocorrem picos de energia para os coeficientes $d_{m,n}$, então, o nível π e os que se encontram em seu entorno formam a série temporal Z , enquanto o restante juntamente com a componente de aproximação formarão X . Não tendo dois picos bem definidos, pode-se optar pelo nível de resolução médio ou algum valor próximo a ele, mediante análise dos gráficos das componentes *wavelet* e escalograma (gráfico com a energia dos coeficientes *wavelet* em cada um dos níveis de resolução), a seção 6.3.2 traz uma aplicação.

Decompondo $\{Y_{t,EW}\}_{t=1}^{T+q}$ em nível p e definido o valor de π , tem-se $\{Z_{t,EW}\}_{t=1}^{T+q} = D_1 + \dots + D_k + \dots + D_\pi$ e $\{x_{t,EW}\}_{t=1}^{T+q} = D_{\pi+1} + \dots + D_p + A_p$, onde D_k é a componente wavelet de detalhes no nível k e A_p a componente de aproximação no nível p . Descartando as q informações em $\{Z_{t,EW}\}_{t=1}^{T+q}$, chega-se a série temporal $\{Z_{t,EW}\}_{t=1}^T$. Para determinar a série temporal $\{X_{t,EW}\}_{t=1}^T$, elimina-se as q informações em $\{x_{t,EW}\}_{t=1}^{T+q}$ e soma-se a média μ , de forma que $\{X_{t,EW}\}_{t=1}^T = \{x_{t,EW}\}_{t=1}^T + \mu$. As séries $\{X_{t,EW}\}_{t=1}^T$ e $\{Z_{t,EW}\}_{t=1}^T$ ($\{y_{t,EW}\}_{t=1}^T = \{X_{t,EW}\}_{t=1}^T + \{Z_{t,EW}\}_{t=1}^T$) são divididas em amostras de treino e validação e submetidas à RNA a fim de se obter a previsão um passo a frente $\hat{y}_{T+1,EW}$, que resulta da soma $\hat{y}_{T+1,EW} = \hat{X}_{T+1,EW} + \hat{Z}_{T+1,EW}$.

A série de resíduos $\{y_{t,R}\}_{t=1}^{T+q}$ (seção 4.1) é submetida a decomposição *wavelet* em nível p , obtendo-se π e consequentemente $\{x_{t,R}\}_{t=1}^{T+q}$ e $\{Z_{t,R}\}_{t=1}^{T+q}$ de forma análoga à descrita para as séries $\{Z_{t,EW}\}_{t=1}^{T+q}$ e $\{x_{t,EW}\}_{t=1}^{T+q}$. Por fim, eliminam-se as q informações em $\{x_{t,R}\}_{t=1}^{T+q}$ e $\{Z_{t,R}\}_{t=1}^{T+q}$ para determinar $\{X_{t,R}\}_{t=1}^T$ e $\{Z_{t,R}\}_{t=1}^T$, de forma que $\{y_{t,R}\}_{t=1}^T = \{X_{t,R}\}_{t=1}^T + \{Z_{t,R}\}_{t=1}^T$, sendo suas previsões usadas no cálculo de $\hat{y}_{T+1,R}$ ($\hat{y}_{T+1,R} = \hat{X}_{T+1,R} + \hat{Z}_{T+1,R}$). Desta forma, chega-se à previsão da série analisada no instante $T + 1$, dada por $\hat{y}_{T+1} = \hat{y}_{T+1,EW} + \hat{y}_{T+1,R}$.

4.3 PREVISÕES PONTUAIS

Uma vez obtidas as séries temporais X e Z , conforme descrito na seção 4.2, num total de quatro séries ($\{\{X_{t,EW}\}_{t=1}^T, \{Z_{t,EW}\}_{t=1}^T, \{X_{t,R}\}_{t=1}^T, \{Z_{t,R}\}_{t=1}^T\}$), estas são modeladas via quatro redes neurais artificiais *feed-forward multilayer perceptron*. Para o treinamento da rede foi utilizado o aplicativo computacional Matlab (8.0), com dez simulações para cada combinação entre janela de entrada de dados ($J=8, \dots, 14$) e número de neurônios na camada escondida ($CE=6, \dots, 12$).

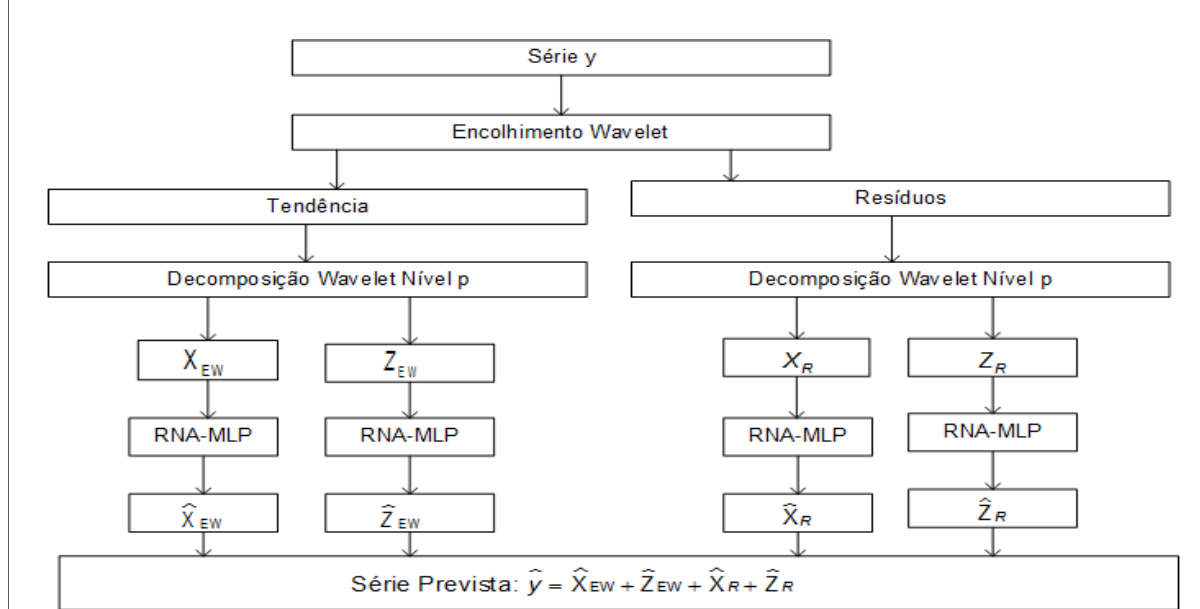
Da modelagem por RNA resultam as previsões representadas nas equações (45) e (46).

$$\hat{y}_{t,EW} = \hat{X}_{t,EW} + \hat{Z}_{t,EW} \quad (45)$$

$$\hat{y}_{t,R} = \hat{X}_{t,R} + \hat{Z}_{t,R} \quad (46)$$

A título de ilustração a Figura 15 traz as etapas do método proposto para a realização das previsões pontuais da série observada.

FIGURA 15 - FLUXOGRAMA COM ETAPAS DO MÉTODO PROPOSTO: PREVISÕES PONTUAIS.



FONTE: O autor (2015).

4.4 PROPOSTA DE INTERVALO DE CONFIANÇA

Tendo previsões pontuais \hat{y}_t da série temporal observada y_t , geram-se B amostras *bootstrap* a partir da amostragem residual sugerida por Tibshirani (1996), revisada na Seção 2.6. Uma vez obtidos os resíduos de treinamento $r_t = y_t - \hat{y}_t$, estes são reamostrados B vezes segundo o procedimento *bootstrap*, sendo o resultado para a b -ésima amostra representado por $r_1^{*b}, r_2^{*b}, \dots, r_t^{*b}$, com $b = 1, \dots, B$. Na sequência, obtêm-se as séries temporais *bootstrapadas* $y_t^{*b} = \hat{y}_t + r_t^{*b}$, com $t = 1, \dots, T$. A fim de se determinar as previsões pontuais das séries temporais *bootstrapadas* y_t^{*b} , utiliza-se a método proposto que se encontra ilustrado na Figura 15 da Seção 4.3.

Para cálculo do erro padrão do t -ésimo valor previsto, Tibshirani (1996) sugere o uso da equação (47).

$$\widehat{se}_{boot}(\hat{y}_t) = \left[\frac{1}{B-1} \sum_{b=1}^B (\hat{y}_t^{*b} - \bar{\hat{y}}_t^*)^2 \right]^{\frac{1}{2}}, \quad (47)$$

onde $\bar{\hat{y}}_t^*$ denota uma média aritmética simples e é dada por $\bar{\hat{y}}_t^* = \frac{1}{B} \sum_{b=1}^B \hat{y}_t^{*b}$, sendo \hat{y}_t^{*b} a previsão de y_t^{*b} . Aqui, propõe-se calcular $\bar{\hat{y}}_t^*$ a partir de uma média ponderada, conforme a Equação (48), cujos pesos δ_b são obtidos através da otimização de um problema de programação não linear, descrito em (49), onde a função objetivo é a raiz do erro quadrático médio entre a combinação linear das previsões \hat{y}_t^{*b} e a série temporal observada y_t , na amostra de treinamento.

$$\bar{\hat{y}}_t^* = \frac{\sum_{b=1}^B \delta_b \hat{y}_t^{*b}}{\sum_{b=1}^B \delta_b} \quad (48)$$

$$\begin{aligned} & \text{Min} \sqrt{\frac{1}{T} \sum_{t=1}^T \left(\sum_{b=1}^B \delta_b \hat{y}_t^{*b} - y_t \right)^2} \\ & \text{s. a.} \quad \delta_b \geq 0 \end{aligned} \quad (49)$$

No modelo de regressão neural $y(x) = f(x, \mathbf{w}) + e(x)$, apresentado na seção 2.6, $f(x, \mathbf{w})$ é a regressão verdadeira e $\hat{y}(x) = g(x, \hat{\mathbf{w}})$, uma aproximação dela. Para uma RNA, a discussão do intervalo de confiança (IC) passa pela distribuição da quantidade $f(x, \mathbf{w}) - \hat{y}(x)$. No caso do uso da técnica *bootstrap*, tem-se B RNAs e passa a ser de interesse a distribuição da quantidade $f(x, \mathbf{w}) - \bar{\hat{y}}^*(x)$, onde $\bar{\hat{y}}^*(x)$ é determinado através de (48). Assumindo que $P(f(x, \mathbf{w}) | \bar{\hat{y}}^*(x))$ e sua inversa $P(\bar{\hat{y}}^*(x) | f(x, \mathbf{w}))$ possuem distribuição gaussiana, o procedimento *bootstrap* permite gerar uma distribuição empírica $P(\hat{y}(x) | \bar{\hat{y}}^*(x))$ que pode estimar $P(\bar{\hat{y}}^*(x) | f(x, \mathbf{w}))$ (HESKES, 1997). Com efeito, a estimativa da variância da distribuição de $P(f(x, \mathbf{w}) | \bar{\hat{y}}^*(x))$ é dada pela equação (50).

$$\hat{\sigma}_g^2(x) = \frac{1}{B-1} \sum_{b=1}^B \left(\hat{y}^{*b}(x) - \bar{\hat{y}}^*(x) \right)^2. \quad (50)$$

Uma vez obtida a estimativa para variância da distribuição $P(f(x, \mathbf{w}) | \bar{\hat{y}}^*(x))$, constrói-se o IC para $f(x, \mathbf{w})$, conforme inequação (51).

$$\bar{\hat{y}}_t^* - t_{1-\frac{\alpha}{2}, gl} \hat{\sigma}_g \leq f(x, \mathbf{w}) \leq \bar{\hat{y}}_t^* + t_{1-\frac{\alpha}{2}, gl} \hat{\sigma}_g, \quad (51)$$

onde $t_{1-\frac{\alpha}{2}, gl}$ provêm da distribuição t de Student com nível de significância α , graus de liberdade gl igual à diferença entre o número de amostras de treinamento e o número de parâmetros do modelo RNA (KHOSRAVI *et al.*, 2014).

Considerando o método proposto (esquematizado na Figura 15), a previsão \hat{y}_t é obtida da modelagem neural das séries temporais X e Z provenientes da decomposição *wavelet* das séries de tendência $y_{t,EW}$ e resíduo $y_{t,R}$. Então, o procedimento descrito nesta seção deve ser aplicado a cada uma das séries temporais X e Z e suas respectivas previsões, somando-se os intervalos de confiança dos regressores dessas séries. Maiores detalhes são mostrados na Seção 6.3.4.

4.5 PREVISÕES INTERVALARES

Para estimar o intervalo de previsão, agrega-se à estimativa da variância do modelo ($\hat{\sigma}_g^2$), obtida por *bootstrap*, a estimativa da variância do ruído ($\hat{\sigma}_e^2$). Assumindo que o ruído é Gaussiano e que sua variância é dependente da entrada x e a regressão neural $y(x) = f(x, w) + e(x)$, a variância do alvo y pode ser escrita na forma da equação (52).

$$[y(x) - \bar{y}_t^*(x)]^2 = [f(x, w) - \bar{y}_t^*(x)]^2 - [e(x)]^2 = \hat{\sigma}_g^2 + \hat{\sigma}_e^2. \quad (52)$$

De acordo com Heskes (1997), pode-se estimar um modelo para σ_e^2 com o ajuste dos resíduos representados na equação (53).

$$R^2(x) = \max\{[y(x) - \bar{y}_t^*(x)]^2 - \hat{\sigma}_g^2, 0\}. \quad (53)$$

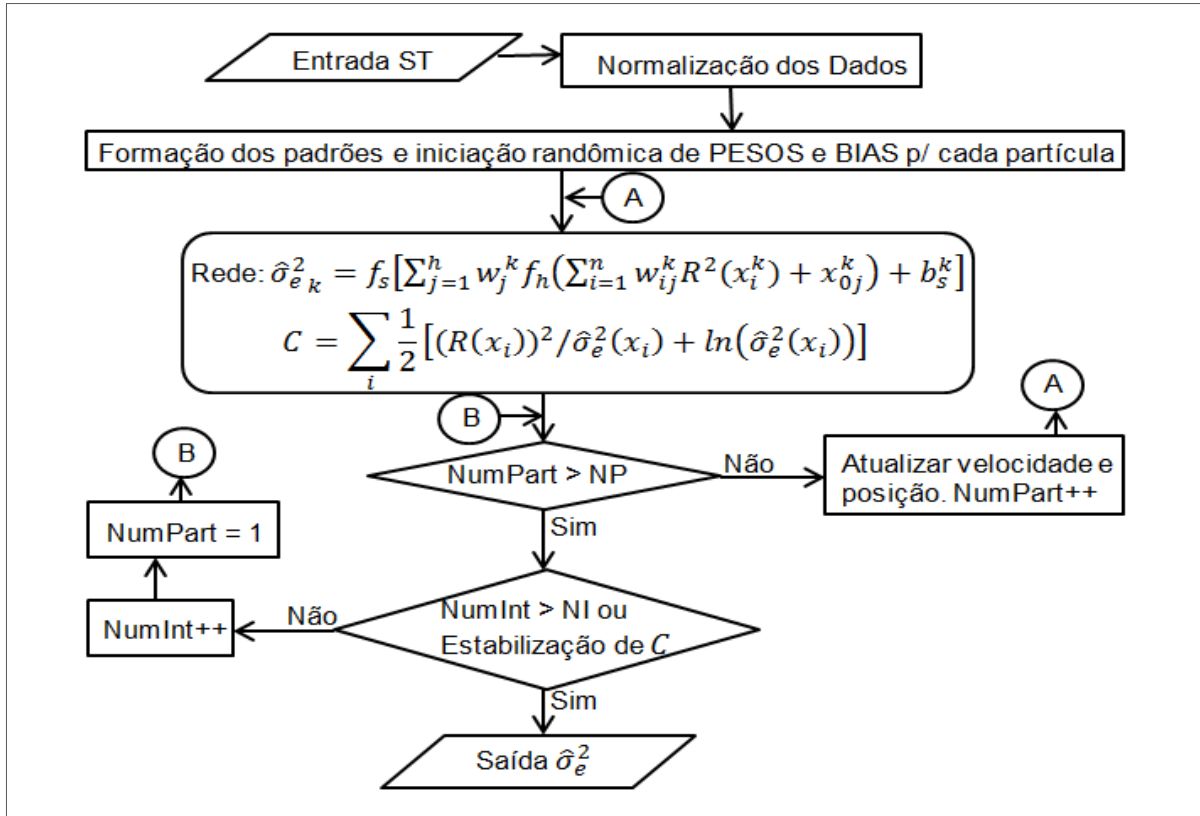
A fim de se ajustar a estatística $R^2(x)$, treina-se uma rede neural artificial com o objetivo de se minimizar uma função custo C , dada na equação 54, adaptada de Nix; Weigend (1995).

$$C = \sum_i \frac{1}{2} \left[\frac{R(x_i)^2}{\hat{\sigma}_e^2(x_i)} + \ln(\hat{\sigma}_e^2(x_i)) \right]. \quad (54)$$

A RNA utilizada é do tipo *feed-forward* e constituída por duas camadas, sendo uma de saída e outra oculta. A otimização de C foi realizada por meio da meta-

heurística PSO (seção 2.7) programada no Matlab. O fluxograma da Figura 16 representa o procedimento utilizado. Nele, NP e NI são iguais ao número de partículas e iterações na meta-heurística PSO, respectivamente. A simbologia usada na RNA está definida na seção 2.3.

FIGURA 16 - FLUXOGRAMA DA REDE-PSO QUE ESTIMA A VARIÂNCIA DO RUÍDO



FONTE: O autor (2015).

Tendo agora uma estimativa para variância do resíduo, pode-se obter o intervalo de previsão a partir de Inequação (55).

$$\hat{y}_t^*(x) - t_{1-\frac{\alpha}{2}, gl} \left(\hat{\sigma}_g(x) + \hat{\sigma}_e(x) \right) \leq y(x) \leq \hat{y}_t^*(x) + t_{1-\frac{\alpha}{2}, gl} \left(\hat{\sigma}_g(x) + \hat{\sigma}_e(x) \right). \quad (55)$$

5 EXPERIMENTO COMPUTACIONAL 1: PREVISÕES PONTUAIS

Neste capítulo são apresentados os resultados obtidos com as previsões das séries temporais *Canadian Lynx*, *Wolf's Sunspot* e *Exchange Rate* pelo método proposto (seções 4.1 – 4.3).

5.1 SÉRIE TEMPORAL CANADIAN LYNX

Os quatro melhores resultados obtidos com o método proposto na previsão da amostra de teste são exibidos na Tabela 2, onde se encontram as bases ortonormais *wavelet* e níveis de decomposição *wavelet* ótimos utilizados no encolhimento, além das estatísticas MSE e MAE. Nas quatro modelagens, procedeu-se o encolhimento *wavelet* com regra de limiarização duro e limiar universal. Para efeitos de comparação será adotada a modelagem 1.

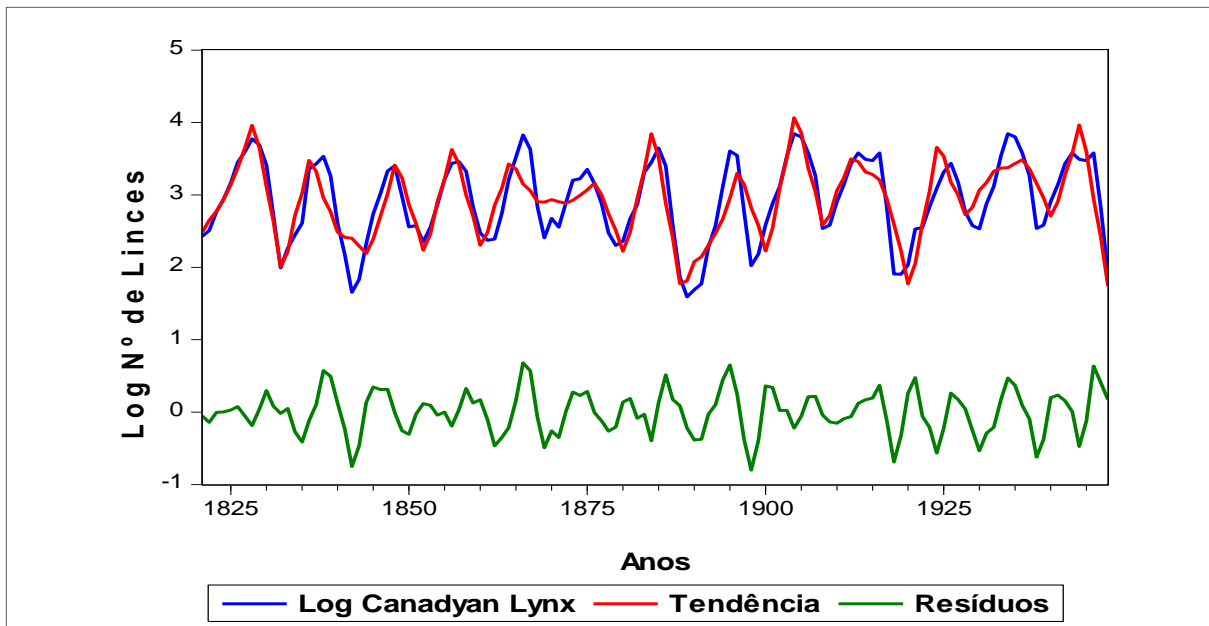
TABELA 2 - BASES WAVELET UTILIZADAS NO ENCOLHIMENTO DA SÉRIE CANADIAN LYNX E ERROS DE PREVISÕES DA AMOSTRA DE TESTE

Modelagem	Base	Nível p	Amostra de Teste	
			MSE	MAE
1	db 3	2	0,00164	0,02803
2	db 2	3	0,00275	0,04962
3	coif 1	2	0,00368	0,06247
4	sym 2	2	0,00291	0,05218

FONTE: O autor (2015).

A série temporal log *Canadian Lynx* com 100 observações (amostra de treino e validação) foi aumentada para 128 instantes. Das 28 novas informações, as 10 primeiras foram obtidas a partir da média das observações em cada um dos instantes do período (tamanho do período = 10 anos). Retrocedendo-se dois períodos, a partir do centésimo termo, tomou-se 18 observações para completar as 128. As séries temporais de resíduos ($y_{t,R}$) e tendência ($y_{t,EW}$) empregadas na modelagem 1 (Tabela 2), juntamente com a série log *Canadian Lynx* (obtida por transformação logarítmica da série *Canadian Lynx*) estão representadas na figura 17. A tabela 3 traz os últimos dez termos da amostra de teste das séries log *Canadian Lynx* (y_t), tendência ($y_{t,EW}$), resíduos ($y_{t,R}$) e suas respectivas previsões ($\hat{y}_t, \hat{y}_{t,EW}$ e $\hat{y}_{t,R}$).

FIGURA 17 - GRÁFICOS DAS SÉRIES TENDÊNCIA E RESÍDUOS ORIUNDOS DO EW DA LOG CANADIAN LYNX



FONTE: O autor (2015).

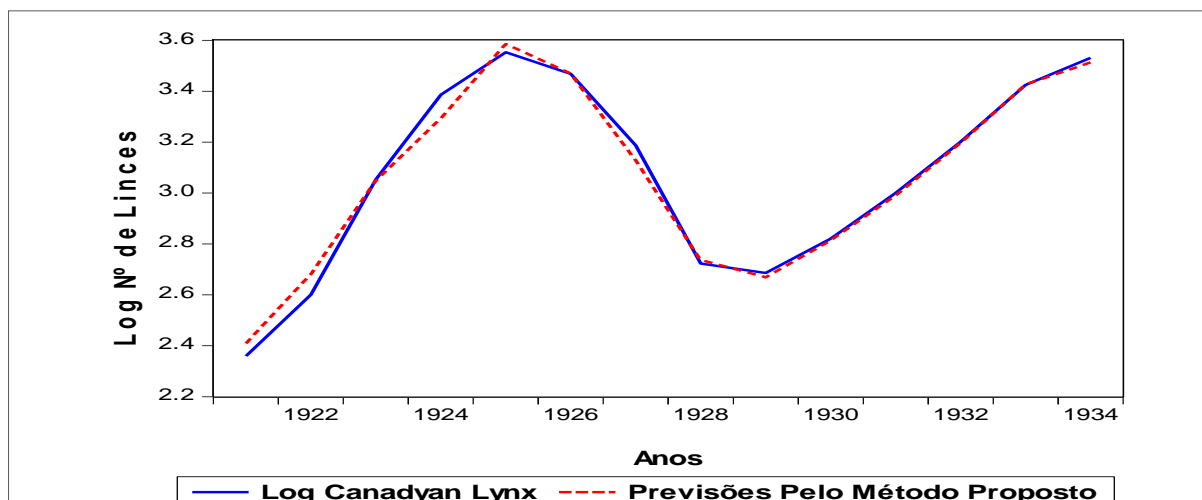
TABELA 3 - PREVISÕES PARA AS 10 ÚLTIMAS OBSERVAÇÕES DA AMOSTRA DE TESTE - CANADIAN LYNX

Ano	y_t	$y_{t,EW}$	$y_{t,R}$	$\hat{y}_{t,EW}$	$\hat{y}_{t,R}$	\hat{y}_t
1925	3,552155	3,444405	0,108749	3,468835	0,116462	3,585297
1926	3,467608	3,720491	-0,252883	3,722204	-0,254050	3,468157
1927	3,186674	3,122128	0,065391	3,073204	0,055319	3,128523
1928	2,723456	2,744809	-0,021354	2,759566	-0,022500	2,737064
1929	2,685742	2,635781	0,049961	2,621671	0,047286	2,668957
1930	2,820858	2,565563	0,255294	2,561876	0,251625	2,813501
1931	3,000000	3,120747	-0,120747	3,103707	-0,114150	2,989553
1932	3,201397	3,481674	-0,280277	3,453862	-0,257890	3,129597
1933	3,424392	3,615404	-0,191012	3,618582	-0,048300	3,425890
1934	3,530968			3,357706	0,155098	3,512804

FONTE: O autor (2015).

Um gráfico comparativo entre y_t e a sua previsão (\hat{y}_t) pelo método proposto é mostrado na Figura 18.

FIGURA 18 - GRÁFICOS DAS SÉRIES LOG CANADIAN LYNX E PREVISÕES PELO MÉTODO PROPOSTO – AMOSTRA DE TESTE.



FONTE: O autor (2015).

Como mencionado anteriormente, os resultados obtidos pelo método proposto na previsão da série temporal *Canadian Lynx* foram comparados a outros trabalhos encontrados na literatura, são eles: Zhang (2003) comparou resultados obtidos com ARIMA, RNA e híbrido ARIMA-RNA; Kajitani, Keith e Mcleod (2005) utilizaram na modelagem da *Canadian Lynx* um modelo RNA *Feed-Forward* (FFNN) e modelo SETAR (*Self-Exciting Threshold Autoregression*); Aladag, Egrioglu e Kadilar, (2009) combinaram RNA recorrentes Elman e ARIMA; Khashei e Bijari, (2010) propuseram modelo híbrido RNA-ARIMA; Ismail e Shabri (2014) usaram Máquinas de Vetores Suporte (SVM) para Regressão e SVM com mínimos quadrados. Os modelos citados neste parágrafo estão na Tabela 4, juntamente com as medidas MSE e MAE. Nesta tabela, a última linha é formada pelos resultados obtidos com o método proposto neste trabalho, as linhas restantes foram obtidas em Ismail e Shabri, (2014).

TABELA 4 - COMPARATIVO MSE E MAE ENTRE MÉTODOS DE PREVISÃO APLICADOS À SÉRIE CANADIAN LYNX

Método	MSE	MAE
ARIMA (Zhang, 2003)	0,0205	0,1123
RNA (Zhang, 2003)	0,0205	0,1121
Híbrido (Zhang, 2003)	0,0172	0,1040
SETAR (Kajitani, 2005)	0,0140	-
RNA (Khashei, 2010)	0,0136	0,0896
FFNN (Kajitani, 2005)	0,0090	-
Híbrido (Aladag, 2009)	0,0090	-
SVR (Ismail, 2014)	0,0085	0,0746
LSSVM (Ismail, 2014)	0,0030	0,0418
Método Proposto	0,0016	0,0281

FONTE: Ismail e Shabri, (2014) – adaptado.

Da Tabela 4, verifica-se que a aplicação do método proposto na realização das previsões da série temporal *Canadian Lynx* obteve redução nas estatísticas de aderência MSE e MAE, em relação a todos os métodos preditivos listados. Comparando-o ao segundo melhor resultado (a saber, o método LSSVM), a redução relativa foi de, aproximadamente, 47% na medida MSE e de 33%, na MAE. Tais resultados mostraram a eficiência preditiva do método proposto diante de abordagens consagradas na Literatura de séries temporais, como: ARIMA, SVM e Híbridos ARIMA-RNA.

5.2 SÉRIE TEMPORAL *WOLF'S SUNSPOT*

A Tabela 5 traz os quatro melhores resultados na modelagem da *Wolf's Sunspot* pelo método proposto.

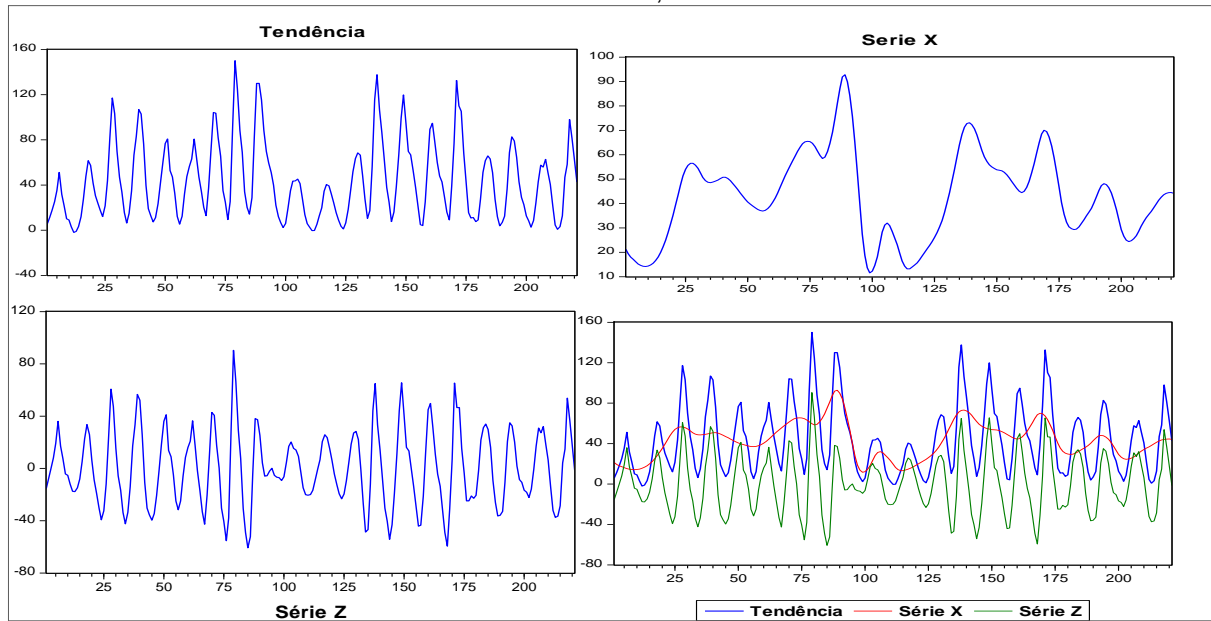
TABELA 5 - MAPE E MAE PARA MODELAGENS ORIUNDAS DE ENCOLHIMENTOS *WAVELET* DISTINTOS- *WOLF'S SUNSPOT*

Modelagem	Base	Nível de decomposição	Limiar	Regra de limiarização	MSE	MAE
1	db 2	3	universal	Suave	243,51	11,68
2	sym 7	2	SURE	Duro	226,79	10,86
3	db 1	3	universal	Duro	251,43	11,95
4	db 10	2	universal	Suave	212,62	10,03

FONTE: O autor (2015).

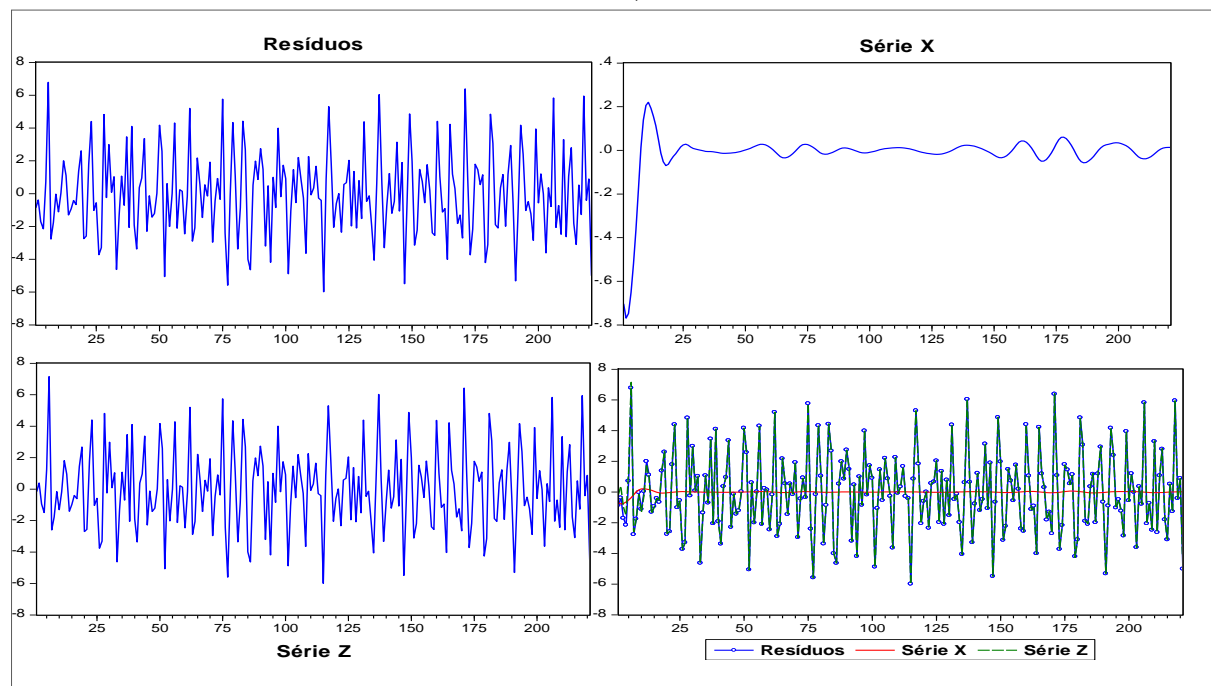
As próximas análises são referentes à modelagem 4 da Tabela 5, pois foi a que mostrou o melhor resultado. Na Figura 19 estão representadas as séries temporais X e Z que resultaram da decomposição ortogonal *wavelet* da série temporal de tendência, com utilização de base *wavelet* da família symlet, mais especificamente a sym10 em nível $p=9$ e $\pi=3$ (nível para construção de X e Z , seção 4.2). As mesmas configurações foram usadas na decomposição *wavelet* da série temporal de resíduos, cujo resultado pode ser observado na Figura 20.

FIGURA 19 - GRÁFICOS DAS SÉRIES: TENDÊNCIA, X E Z - WOLF'S SUNSPOT.



FONTE: O autor (2015).

FIGURA 20 - GRÁFICOS DAS SÉRIES: RESÍDUOS, X E Z - WOLF'S SUNSPOT.



FONTE: O autor (2015).

A título de ilustração, na Tabela 6 estão listadas as 10 últimas observações (y_t) da amostra de teste, juntamente com as suas previsões (\hat{y}_t) que foram obtidas com a soma das previsões da tendência ($\hat{y}_{t,EW}$) e resíduo *wavelet* ($\hat{y}_{t,R}$).

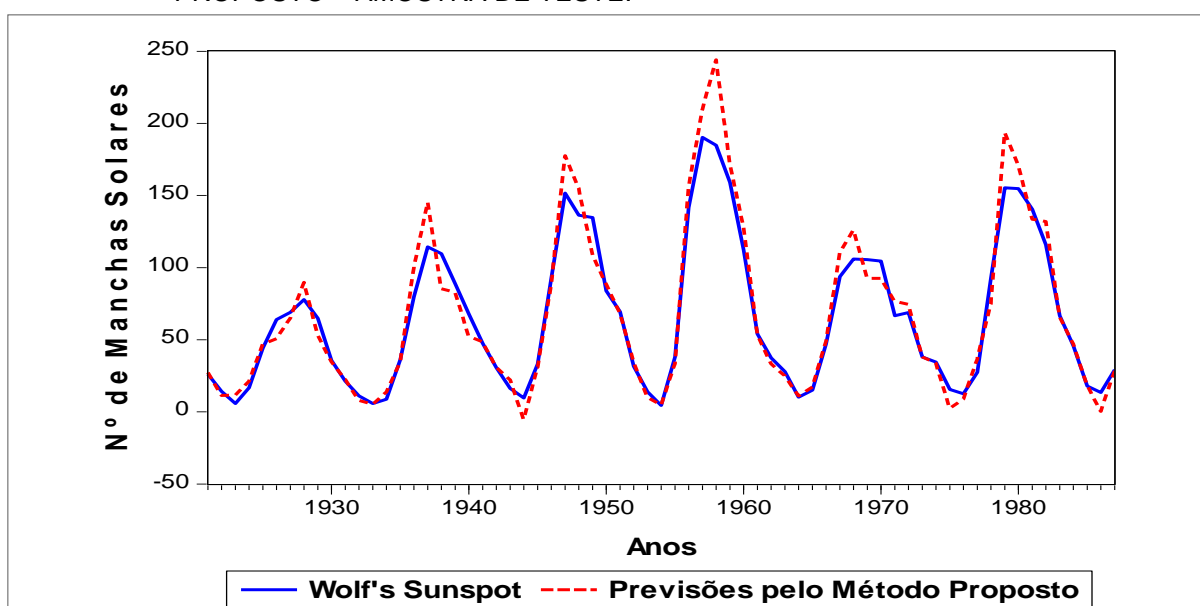
TABELA 6 - PREVISÕES PARA AS 10 ÚLTIMAS OBSERVAÇÕES DA AMOSTRA DE TESTE -
WOLF'S SUNSPOT

Ano	y_t	$y_{t,EW}$	$y_{t,R}$	$\hat{y}_{t,EW}$	$\hat{y}_{t,R}$	\hat{y}_t
1978	92,5	93,627	-1,126	75,128	-0,289	74,838
1979	155,4	151,792	3,607	188,089	5,568	193,658
1980	154,7	154,364	0,336	170,555	0,407	170,962
1981	140,5	137,172	3,327	104,127	1,327	105,454
1982	115,9	115,419	0,480	131,166	0,662	131,828
1983	66,6	71,076	-4,476	69,691	-3,904	65,787
1984	45,9	43,558	2,342	45,754	1,261	47,016
1985	17,9	19,679	-1,779	20,206	-1,727	18,479
1986	13,4	16,480	-3,080	5,861	-5,550	0,311
1987	29,2			27,213	1,263	28,476

FONTE: O autor (2015).

Na Figura 21 estão representados os gráficos das observações de *Wolf's Sunspot* e de suas respectivas previsões oriundas do método proposto.

FIGURA 21 - GRÁFICOS DAS SÉRIES WOLF'S SUNSPOT E PREVISÕES PELO MÉTODO PROPOSTO – AMOSTRA DE TESTE.



FONTE: O autor (2015).

Na Tabela 7 podem ser observados os resultados de previsões da *Wolf's Sunspot* obtidos por outros autores e publicados na literatura especializada e, os alcançados por meio do método proposto.

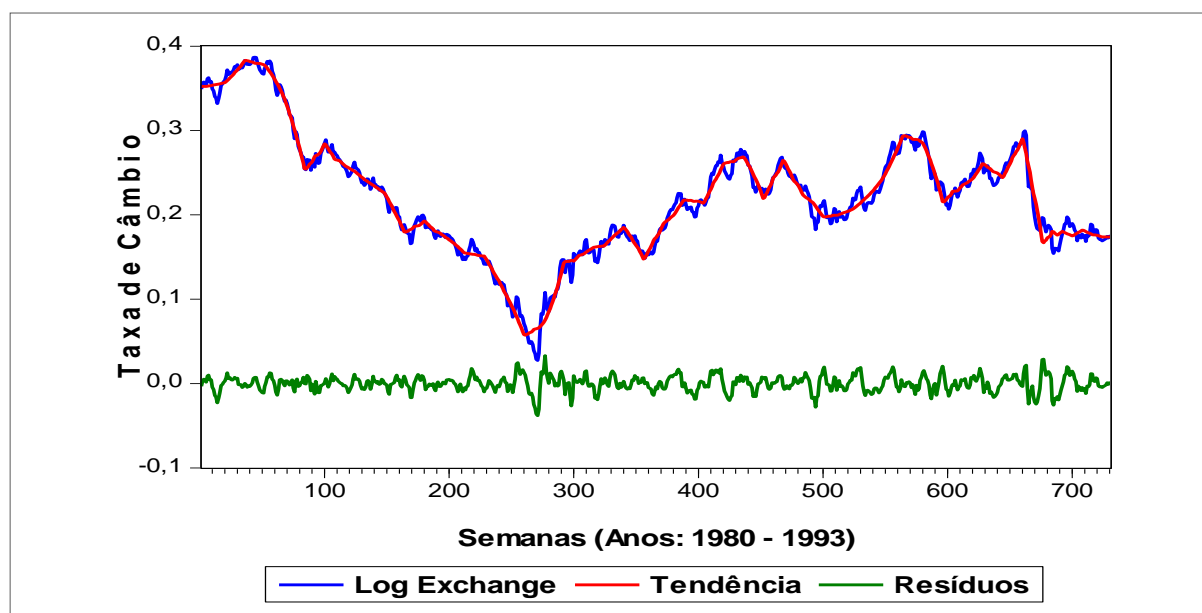
TABELA 7 - COMPARATIVO MSE E MAE ENTRE MÉTODOS DE PREVISÃO APLICADOS À SÉRIE *WOLF'S SUNSPOT*

Método	MSE	MAE
ARIMA (Zhang, 2003)	306,082	13,0337
RNA (Zhang, 2003)	351,194	13,5444
Híbrido (Zhang, 2003)	280,160	12,7802
ARIMA (Adhikari, 2012)	483,500	17,6300
RNA (Adhikari, 2012)	494,900	15,5800
EANN (Adhikari, 2012)	492,900	14,7100
Ensemble (Adhikari, 2012)	247,700	12,5000
RNA (Khashei, 2011)	234,206	12,1180
Híbrido (Khashei, 2011)	218,642	11,4470
ARIMA/PNN(Khashei, 2012)	234,775	11,5485
Método Proposto	212,618	10,0301

FONTE: O autor (2015).

5.3 SÉRIE TEMPORAL *EXCHANGE RATE*

Entre as várias configurações de EW utilizadas na modelagem da série transformada Log *Exchange Rate*, a que apresentou menor erro de previsão era constituída por uma base *wavelet* daubechies de ordem três, decomposta em nível quatro, limiar universal e regra de limiarização duro. Na figura 22 estão representadas as séries: logaritmo de *Exchange Rate* (y_t), a tendência ($y_{t,EW}$) e resíduos ($y_{t,R}$). Na Tabela 8 estão registradas as 10 últimas previsões (amostra de teste) para estas séries.

FIGURA 22 - GRÁFICOS DAS SÉRIES LOG *EXCHANGE RATE*, TENDÊNCIA E RESÍDUOS

FONTE: O autor (2015).

TABELA 8 - PREVISÕES PARA AS 10 ÚLTIMAS OBSERVAÇÕES DA AMOSTRA DE TESTE -
EXCHANGE RATE

Data	y_t	$y_{t,EW}$	$y_{t,R}$	$\hat{y}_{t,EW}$	$\hat{y}_{t,R}$	\hat{y}_t
29/10/93	0,17085	0,17524	0,00440	0,17827	-0,00562	0,17265
05/11/93	0,17079	0,17514	-0,00435	0,17404	-0,00351	0,17054
12/11/93	0,16923	0,17483	-0,00560	0,17688	-0,00570	0,17118
19/11/93	0,17020	0,17445	-0,00425	0,17593	-0,00504	0,17089
26/11/93	0,17094	0,17410	-0,00317	0,17567	-0,00309	0,17258
03/12/93	0,17179	0,17370	-0,00192	0,17515	-0,00264	0,17251
10/12/93	0,17458	0,17339	0,00012	0,17524	-0,00184	0,17340
17/12/93	0,17266	0,17310	-0,00044	0,17557	-0,00021	0,17536
24/12/93	0,17470	0,17275	0,00194	0,17458	-0,00169	0,17289
31/12/93	0,17351			0,17408	-0,00119	0,17527

FONTE: O autor (2015).

Na Tabela 9 encontra-se um comparativo envolvendo os resultados obtidos por Zhang (2003) e Khashei e Bijari (2010, 2011, 2012). Novamente, o método proposto se apresenta com os menores erros de previsão e com ganhos preditivos relevantes quando comparados aos resultados alcançados pelos autores citados. A Figura 23 ilustra a proximidade da Log *Exchange Rate* com sua previsão (amostra de teste) pelo método proposto.

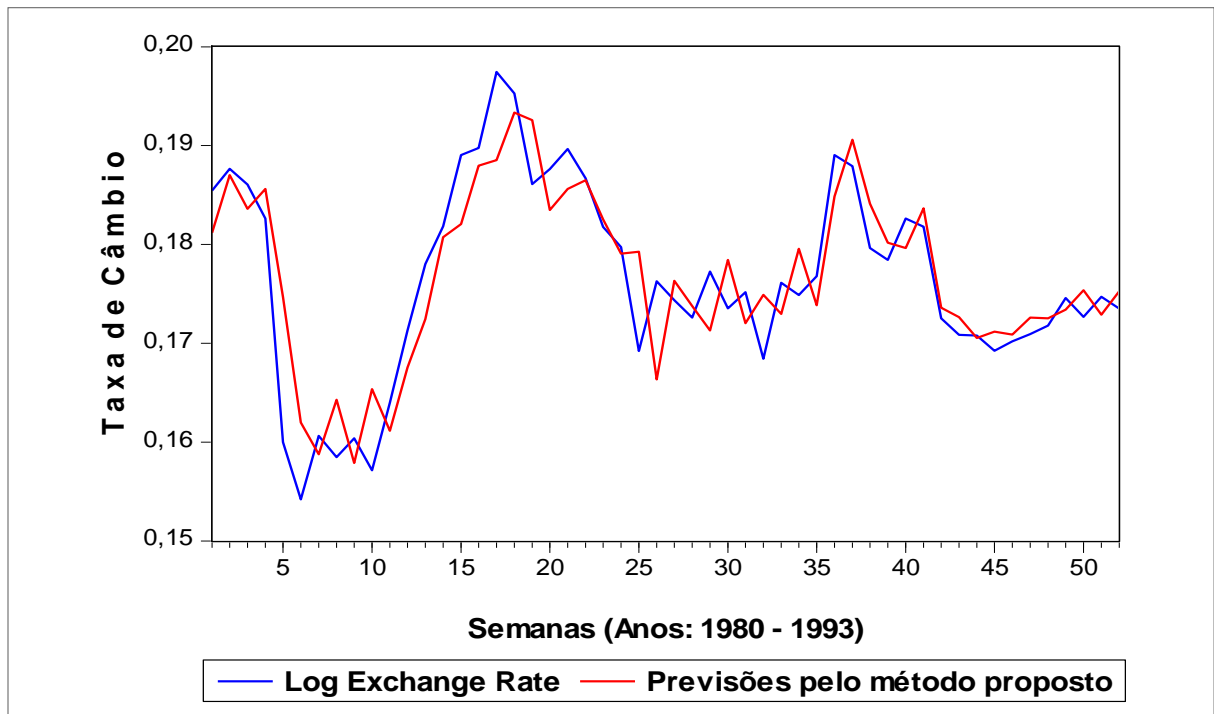
TABELA 9 - COMPARATIVO MSE E MAE ENTRE MÉTODOS DE PREVISÃO APLICADOS À SÉRIE
LOG EXCHANGE RATE

Método	MSE	MAE
ARIMA (Zhang, 2003)	4,52977	0,005360
RNA (Zhang, 2003)	4,52657	0,005251
Híbrido (Zhang, 2003)	4,35907	0,005121
RNA (Khashei, 2010)	3,76399	0,005107
RNA (Khashei, 2011)	3,64724	0,004969
ARIMA/PNN (Khashei, 2012)	3,35623	0,004717
RNA/PNN (Khashei, 2012)	2,77122	0,004271
Método Proposto	2,17449	0,003620

FONTE: O autor (2015)

NOTA: multiplicar todos os valores de MSE por 10^{-5} .

FIGURA 23 - GRÁFICOS DAS SÉRIES LOG *EXCHANGE RATE* E PREVISÕES PELO MÉTODO PROPOSTO – AMOSTRA DE TESTE.



FONTE: O autor (2015).

6 EXPERIMENTO COMPUTACIONAL 2: PREVISÕES INTERVALARES

Neste capítulo, são mostrados os resultados empíricos da aplicação do método proposto (seções 4.1 – 4.5) em toda a sua extensão, incluindo a estimativa dos intervalos de confiança e previsão, para tanto foi escolhida a série de vazões médias mensais formada a partir de informações coletadas no posto 266-Itaipu. Os resultados obtidos com a modelagem da série de vazões pelo método proposto foram comparados com outras abordagens já consagradas na literatura, tais como os individuais Box e Jenkins (modelos ARIMA) e Redes Neurais Artificiais e os compostos *Wavelet*-ARIMA e *Wavelet*-RNA. Foram feitas previsões de um passo à frente, em um horizonte de 49 meses. Para a avaliação dos métodos, foram utilizados os erros MAPE (Erro Médio Percentual Absoluto) e MAE (Erro Médio absoluto). O capítulo inicia descrevendo o problema das previsões de vazões, na sequência são apresentadas as previsões obtidas pelos métodos individuais (ARIMA e RNA) e compostos (*Wavelet*-ARIMA e *Wavelet*-RNA), terminando com o método proposto.

6.1 O PROBLEMA DA PREVISÃO DE VAZÕES

Existem vários fatores que podem influenciar a vazão em um determinado ponto de um curso de água, entre eles: evaporação, precipitação, infiltração e percolação. Sendo a precipitação a principal componente na formação da vazão e responsável pela característica aleatória do processo hidrológico (GUILHON, 2002). Para o estudo de vazões em usina hidrelétricas, podem-se destacar as vazões: natural, incremental, turbinada e vertida. A vazão natural é a que é observada em um posto de medição considerando o rio na sua condição natural. A vazão incremental, por sua vez, é toda vazão lateral captada entre um posto de medição e todos os postos de medição imediatamente a montante deste. A vazão turbinada é aquela que passa pelas turbinas e gera energia, enquanto a vazão vertida é aquela que passa pelos vertedouros da usina e não gera energia.

A fim de se modelar a realidade hídrica são formadas séries temporais de vazões médias diárias, semanais e mensais, o Operador Nacional do Sistema Elétrico (ONS) disponibiliza um banco de dados contendo o registro histórico de vazões naturais de todas as usinas hidrelétricas que compõem o Sistema Interligado

Nacional (SIN). O primeiro registro de vazão natural média mensal é referente ao mês de janeiro de 1931, sendo o acesso aos registros associados ao código denominado posto. No caso da Itaipu o posto artificial tem o número 66 e o natural 266.

Estimativas da Empresa de Pesquisa Energética (EPE) apontam que o consumo de energia elétrica no Brasil em 2020 será 650.092 GWh, sendo predominante a produção hidráulica em virtude da situação privilegiada do país em relação aos recursos hídricos. A produção de energia a partir das usinas hidrelétricas tem ligação direta com a quantidade de água (vazão afluente) disponível para a movimentação das turbinas geradoras. De forma que longos períodos de seca podem levar a uma diminuição da capacidade geradora de energia da usina. A relação entre vazões afluentes e geração de energia implica em um cuidado especial com os dados que cercam o problema.

No Brasil, o *Operador Nacional do Sistema Elétrico* (ONS) utiliza as informações sobre as vazões na otimização da geração de energia e elaboração de estratégias que minimizem efeitos negativos de eventuais enchentes ou secas. Para tanto, o ONS faz previsão de vazões naturais para locais de aproveitamento hidrelétrico que constituem o Sistema Interligado Nacional (SIN), sendo o PREVIVAZ o modelo de previsão mais utilizado.

O PREVIVAZ, usado na previsão de vazões e desenvolvido pelo Centro de Pesquisas de Energia Elétrica (CEPEL), é um modelo estocástico univariado (Maceira *et al.*, 1999), este modelo é usado para prever as vazões semanais da maioria das bacias do SIN. O PREVIVAZ usa modelos lineares do tipo PARMA(p,q) (Box e Jenkins, 1970), além de transformações de séries históricas, tipo Box-Cox ou Logarítmicas. Relatório emitido pelo ONS para previsões de vazões (dezembro de 2011) menciona outros modelos além do PREVIVAZ. Seguem alguns exemplos de outros modelos e as bacias em que foram aplicados:

- Para a bacia do rio Uruguai foi utilizado o modelo MPCV – Modelo de Previsão de Classes de Vazões;
- Para a bacia do rio Iguaçu, foi utilizado o modelo de Redes Neurais Fuzzy Recorrentes;
- Para o trecho da bacia do rio Paraná entre as UHEs Rosana, Porto Primavera e Itaipu, foi utilizado o modelo SMAP-MEL.

Entre os métodos usados pelo ONS, já citados, ressalta-se a frequência de uso dos métodos de Box e Jenkins (Box *et al.* 1994), devido às suas estruturas matemáticas relativamente simples. Os modelos Box-Jenkins para previsão de séries temporais pertencem à família dos *autorregressivos médias móveis* (ARMA), que é subdividido em dois outros modelos: o *autorregressivo* (AR) e *médias móveis* (MA). O segundo (MA) tem processos de ordem q em que cada observação é gerada por uma média ponderada dos erros aleatórios q períodos no passado. O método $AR(p)$ modela uma autorregressão da variável observável com ela mesma, defasada para os p períodos de defasagem em que a autocorrelação parcial entre as variáveis é significativa. Quando a série temporal é não estacionária, sugere-se que seja feita a diferenciação. Neste caso, o modelo em questão pode ser o ARIMA (p,d,q) , onde d representa a ordem de diferenciação (MORETTIN, 2006).

Na Itaipu a previsão das vazões diárias é realizada em duas fases: uma determinística e outra estocástica. Na fase determinística usa-se o método de Muskingum para a propagação das vazões. Na fase estocástica utiliza-se a metodologia Box e Jenkins. A série temporal supracitada é analisada a fim de se determinar os parâmetros do modelo ARIMA que será utilizado na previsão de vazões médias diárias dez dias à frente. A calibração dos parâmetros do modelo ARIMA escolhido foi realizada no início das operações da Itaipu na década de 80. Entre os principais formadores do lago de Itaipu, estão os rios Ivai e Piquiri, localizados na margem esquerda do rio Paraná. Para estes rios, o departamento especializado da Itaipu faz estimativas da vazão mensal.

Segundo WANG *et al.* (2006a), os processos de geração de vazões são sazonais e não lineares. Os mecanismos não lineares são mais fortes em escalas menores de tempo, sendo as vazões diárias e mensais não lineares (Wang, *et al.*, 2006b). São vários os trabalhos existentes na literatura que discutem a previsão de séries oriundas de processos não lineares, sendo predominante para este caso o uso de Redes Neurais Artificiais (RNA). Ballini (2000); Wang *et al.*, (2006a); Valenzuela *et al.* (2008); Ferreira (2012); Abreu, *et al.* (2012); Teixeira, Jr *et al.* (2012a), entre outros, utilizaram Redes Neurais Artificiais (RNA) na previsão de séries temporais.

Embora as RNAs possam ser aplicadas em problemas lineares e não lineares e como ferramenta única em previsões de séries temporais, uma breve inspeção na literatura especializada detecta o uso rotineiro de combinações de previsores e

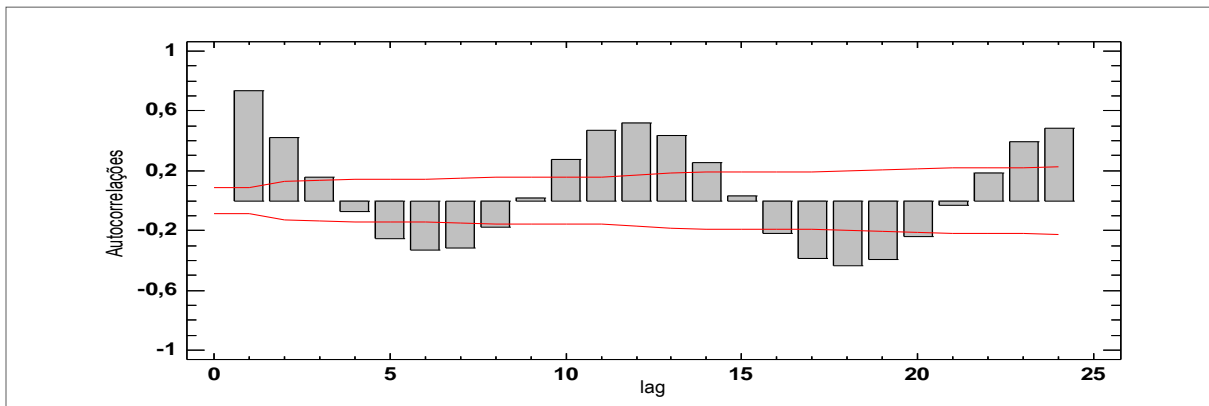
previsões. Na associação de previsores pode-se destacar a combinação dos previsores ARIMA e RNA, ARIMA e Support Vector Machine (SVM) e ARIMA e GARCH, nestes casos os autores partem do pressuposto que as séries são geradas por processos lineares e não lineares, veja Zhang (2003); Hwarng e Ang (2001); Valenzuela *et al.* (2008); Abreu, *et al.* (2012); Sujjavirivasup (2013); Yaziz, *et al.* (2013). Outros usam fazer um pré-processamento da série a partir de *Wavelet*, Filtros de Kalman, Filtros de Partículas e *Singular Spectrum Analysis* (SSA), para então alimentar o previsor, é o caso de Lima (2011); Menezes, *et al.* (2012); Hirpa (2013); Souza (2013). Com relação à combinação de previsões, Bates e Granger (1969) apresentaram trabalho que mostra relativa eficiência desta técnica.

A previsão de afluências nos reservatórios das usinas hidrelétricas é de grande importância no planejamento da produção de energia, propicia ao gestor do sistema elementos para a otimização do total da energia produzida, explorando toda a sua potencialidade. Também faz parte das ações da Itaipu o monitoramento do nível do rio Paraná à jusante da barragem, por conta de tratados firmados com Argentina e a preocupação com o fenômeno enchente sobre as populações ribeirinhas. Desta forma, as previsões de afluências realizadas com certa antecedência e precisão permitem aos tomadores de decisões elaborarem estratégias que minimizem os efeitos negativos de cheias e secas extraordinárias. As previsões de afluências ainda fornecem informações importantes para outros setores envolvidos com o reservatório, tais como: abastecimento urbano e agrícola, piscicultura, aquicultura, recreação e lazer.

6.2 MÉTODOS: ARIMA, RNA, *WAVELET*-ARIMA E *WAVELET*-RNA

A análise do correlograma (Figura 24) da série temporal de vazões mensais observada mostra a existência de componente de sazonalidade com período igual a doze e autocorrelações não decrescentes nos lags múltiplos de doze, indicando não-estacionariedade. A indicação de não-estacionariedade foi reforçada pelo teste de Dickey Fuller aumentado para um nível de significância de 5% (p-valor obtido = 0,06).

FIGURA 24 - CORRELOGRAMA DA SÉRIE DE VAZÕES MENSAIS



FONTE: O autor (2015).

As séries temporais de vazões possuem como característica a não-estacionariedade, com a presença de componente sazonal. Ballini (2000) sugere extrair tais componentes do processo de previsão através de um tratamento estatístico, onde a série original é transformada em outra a partir de uma padronização. Quando as séries estudadas apresentam indícios de não-estacionariedade, Box; Jenkins e Reinsel (1994) sugerem a diferenciação da série a fim de se obter séries estacionárias, neste caso o modelo escolhido é o $ARIMA(p, d, q)$, onde d é a ordem de diferenciação da série. Como a série observada $(y_t)_{t=1}^{492}$ possui componente sazonal, foi necessário o uso do modelo ARIMA multiplicativo, representado $ARIMA(p, d, q) \times (P, D, Q)_s$. Sendo S o período sazonal, D a ordem do operador diferença sazonal, QS a ordem do modelo médias móveis sazonal e PS a ordem do modelo autorregressivo sazonal (SOUZA; CAMARGO, 1996).

A fim de eliminar o efeito de não-estacionariedade e sazonalidade foi aplicada uma diferenciação de ordem $D = 12$. Com a série diferenciada, fez-se a análise dos gráficos da FAC e FACP, podendo assim definir possíveis modelos para o ajuste dos dados. Testes com os resíduos indicaram que entre os modelos candidatos o $ARIMA(1,0,2) \times (0,1,2)_{12}$ é o que melhor se ajusta aos dados da série de vazões em estudo.

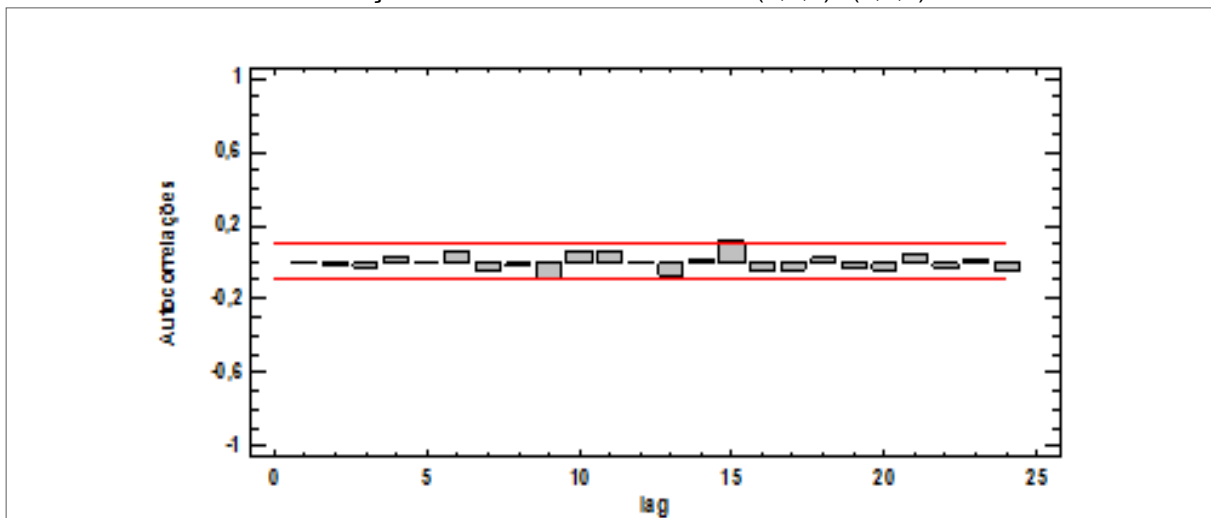
Para a validação do modelo ARIMA ajustado à série observada foram analisados os resíduos com a construção de gráficos de autocorrelações, buscando

mostrar que os mesmos são não autocorrelacionados. A validação é reforçada com o periodograma integrado, segundo Morettin e Toloi (2006) se o modelo não for adequado, haverá desvios sistemáticos da reta no gráfico do periodograma acumulado $C(f_j)$ pela frequência (f_j) .

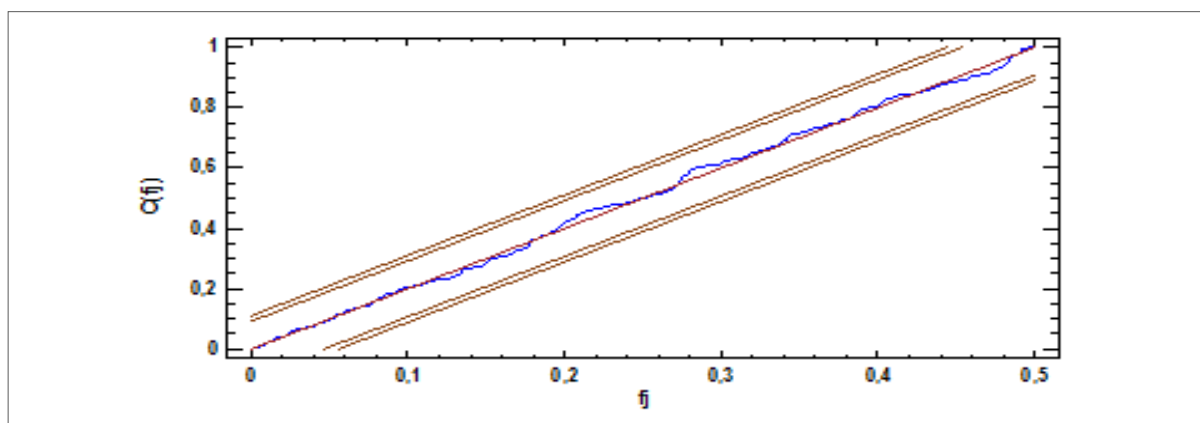
A Figura 25 mostra o correlograma dos resíduos provenientes da modelagem. Não considerando a autocorrelação no lag 15, pode-se afirmar pelo teste da autocorrelação residual que o modelo ajustado é adequado. Visto que o correlograma para a série de resíduos mostra valores situados num intervalo no qual a autocorrelação é considerada nula.

O teste de Box-Pierce corrobora com a conclusão obtida com o teste da autocorrelação residual. Como o p-valor (0,203) obtido com o teste de Box-Pierce é maior que 0,05, não se pode rejeitar a hipótese que a série de resíduos é randômica ao nível de 5%. O periodograma (figura 26) reforça os testes anteriores, confirmando que o modelo $ARIMA(1,0,2) \times (0,1,2)_{12}$ é adequado aos dados, não tendo nos resíduos parte sistemática a ser modelada, estando os resíduos próximos ao denominado ruído branco.

FIGURA 25 - AUTOCORRELAÇÕES DOS RESÍDUOS: $ARIMA(1,0,2) \times (0,1,2)_{12}$



FONTE: O autor (2015).

FIGURA 26 - PERIODOGRAMA PARA OS RESÍDUOS: ARIMA(1,2,2)X(0,1,2)¹²

FONTE: O autor (2015).

A busca pela RNA que melhor ajustasse os dados identificou a RNA-MLP com tamanho de janela de entrada igual a quatro ($J=4$) e camada escondida composta por cinco neurônios artificiais ($CE=5$), conforme Tabela 10, onde estão listadas seis das vinte estruturas testadas. Nesta tabela podem ser observados o Erro Médio Percentual Absoluto (MAPE), Erro Percentual Absoluto (APE) mínimo e o APE máximo para as seis configurações listadas.

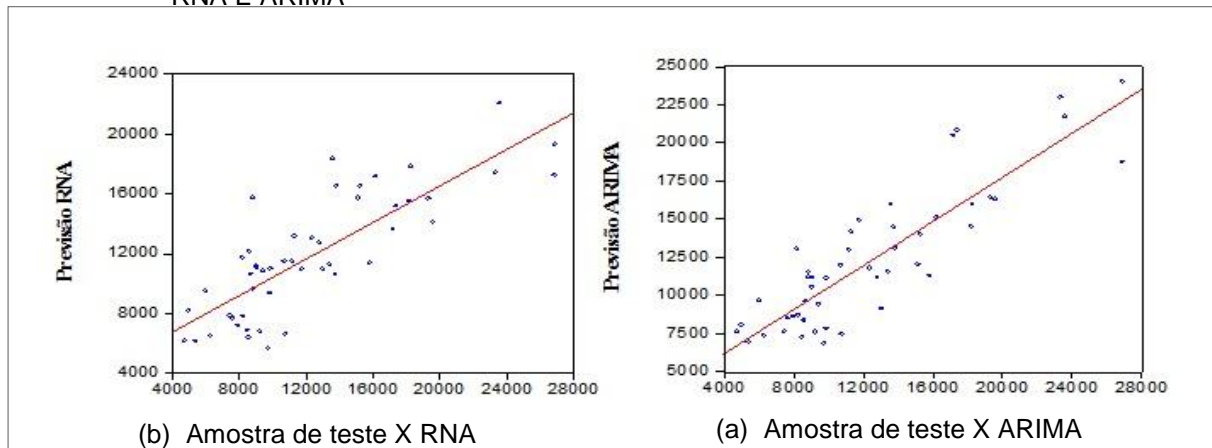
TABELA 10 - APE MÍNIMO, MÁXIMO E MAPE PARA SEIS ESTRUTURAS DISTINTAS DE RNA-MLP

J	CE	APE Mínimo			APE Máximo			MAPE		
		Treino	Validação	Teste	Treino	Validação	Teste	Treino	Validação	Teste
5	5	0,060	0,20	0,23	112,3	76,7	79,4	17,8	18,4	21,3
4	4	0,040	0,67	0,67	113,5	70,3	86,4	19,3	20,1	21,6
5	6	0,020	0,44	0,41	132,9	92,5	159,5	16,8	18,6	26,3
5	7	0,005	0,07	0,49	109,3	72,1	146,7	17,2	18,8	27,2
4	5	0,020	0,04	0,68	122,1	87,4	49,4	18,2	18,3	18,9
4	6	0,004	0,07	1,22	126,8	88,3	171,2	18,3	18,9	24,1

FONTE: O autor (2015).

A amostra de teste, formada pelas últimas 49 observações, apresentou MAPE igual a 18,01% com o modelo ARIMA e 18,9% com a modelagem por RNA. Nas Figuras 27a, 27b estão os gráficos de dispersão da amostra de teste versus previsão por RNA e ARIMA, respectivamente. Os gráficos em questão são um indicativo que se pode melhorar a qualidade das previsões obtidas com os métodos individuais ARIMA e RNA. Na Figura 27a o coeficiente de correlação encontrado foi de 0,84 e na Figura 27b obteve-se coeficiente igual a 0,87.

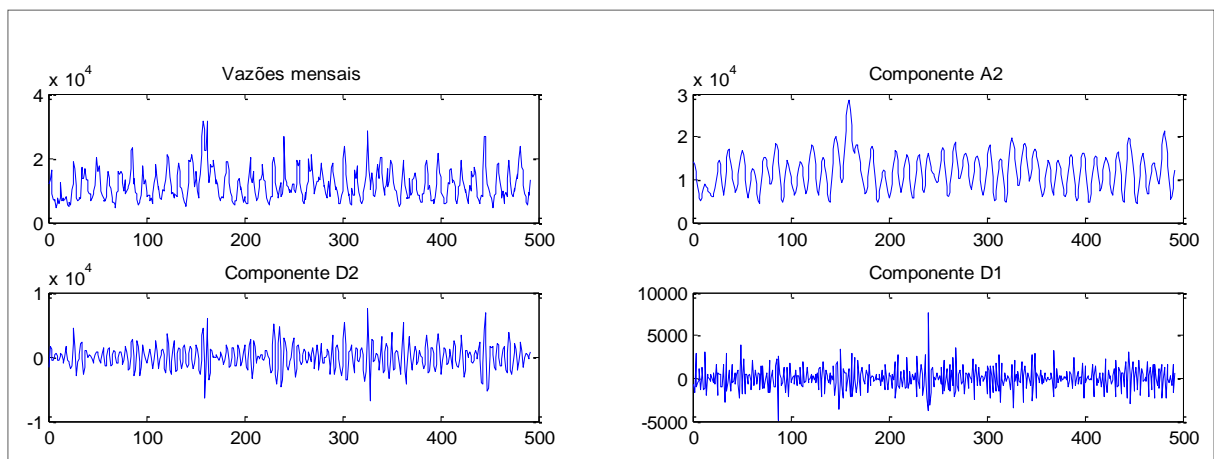
FIGURA 27 - DIAGRAMAS DE DISPERSÃO ENTRE A AMOSTRA DE TESTE E PREVISÕES POR RNA E ARIMA



FONTE: O autor (2015).

Os métodos compostos *Wavelet*-ARIMA e *Wavelet*-RNA iniciam pela decomposição da série temporal observada. Na figura 28 podem-se observar os gráficos da série de vazões mensais e as componentes *wavelets* de aproximação (A2) e detalhes (D1 e D2) obtidas com a base ortonormal *wavelet* de Daubechies com momento 10 (db 10).

FIGURA 28 - GRÁFICO DA SÉRIE DE VAZÕES E COMPONENTES WAVELET



FONTE: O autor (2015).

O Quadro 2 traz um resumo dos modelos autorregressivos (AR), médias móveis (MA) e configurações das RNA utilizadas na modelagem das componentes *wavelet* A2, D1 e D2, sendo CE a quantidade de neurônios da camada escondida com função de ativação tangente hiperbólica (tansig) e J o tamanho da janela de entrada de dados. A camada de saída possui um único neurônio e a função de ativação é linear (purelin).

QUADRO 2 - CONFIGURAÇÕES RNA E ORDENS DOS MODELOS AR E MA PARA A2, D1 E D2.

Componente de aproximação de nível 2 (A2):	
RNA_1: <i>feedforward</i> , J=20, CE=9, <i>tansig</i> (camada escondida), <i>purelin</i> (camada de saída).	
AR(1), AR(2), AR(4), AR(12), AR(13), AR(24), MA(1 até 3), MA(13) e MA(24)	
Componente de detalhe de nível 1 (D1):	
RNA_2: <i>feedforward</i> , J=20, CE=5, <i>tansig</i> (camada escondida), <i>purelin</i> (camada de saída)	
AR(1 até 8), MA(1 até 8)	
Componente de detalhe de nível 2 (D2):	
RNA_3: <i>feedforward</i> , J=21, CE=9, <i>tansig</i> (camada escondida), <i>purelin</i> (camada de saída)	
AR(1 até 8), AR(10), AR(11), AR(24), MA(1 até 4), MA(6), MA(7) e MA(10)	

FONTE: O autor (2015).

As previsões das componentes wavelet obtidas por ARIMA e RNA foram combinadas linearmente a fim de se determinar as previsões da série de vazões.

O MAPE para a amostra de teste obtido com a aplicação do método *Wavelet*-ARIMA foi de 15,20% e 14,60% com a utilização do *Wavelet*-RNA, mostrando a eficiência desses métodos em relação aos individuais RNA e ARIMA.

6.3 MÉTODO PROPOSTO: RESULTADOS EMPÍRICOS

Nesta seção estão detalhadas as etapas de execução do método proposto (capítulo 4) até a previsão intervalar, além de comparativo com os resultados dos métodos da seção 6.2.

Foram realizadas quinze modelagens distintas, utilizando-se mais de um tipo de base *wavelet* (daubechies, symlets e coiflets), níveis de decomposição, limiar (universal e SURE) e regra de limiarização (suave e duro). Na Tabela 11 estão listados os cinco melhores resultados para a amostra de teste.

TABELA 11 - MAPE E MAE PARA MODELAGENS ORIUNDAS DE ENCOLHIMENTOS WAVELET DISTINTOS- SÉRIE DE VAZÕES

Modelagem	Base	Nível de decomposição	Limiar	Regra de limiarização	MAPE	MAE
1	haar	2	Universal	duro	9,56	1132,97
2	haar	1	SURE	duro	11,34	1347,73
3	haar	2	SURE	suave	10,62	1245,04
4	haar	4	Universal	suave	10,82	1281,38
5	db 2	2	SURE	suave	11,79	1403,90

FONTE: O autor (2015).

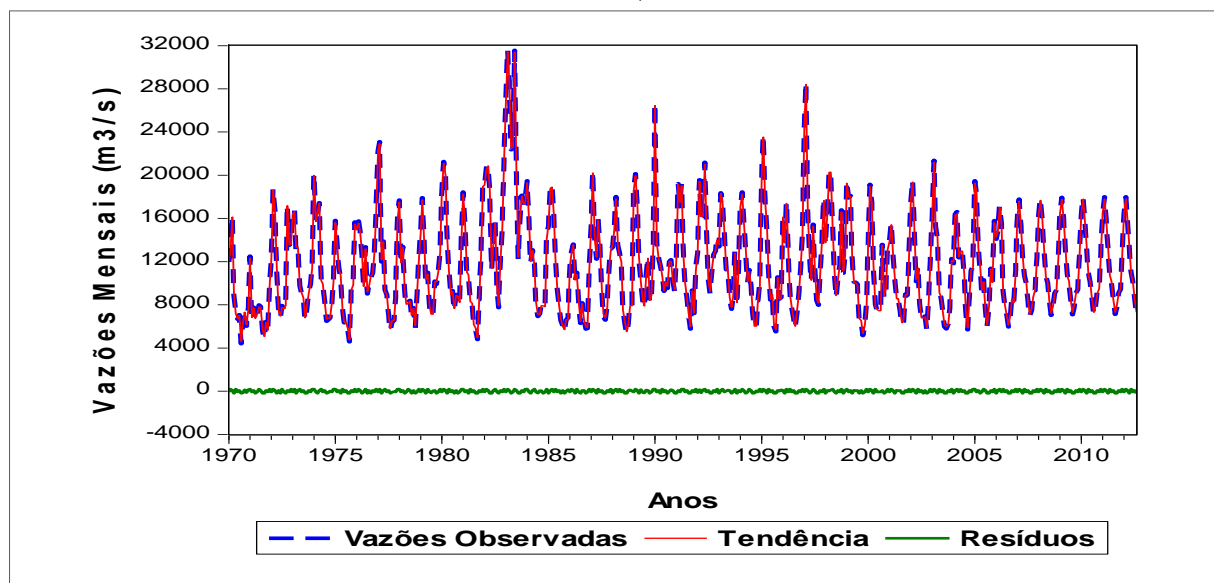
As discussões nas seções 6.3.1 até 6.3.5 são concernentes ao melhor resultado (modelagem 1) expresso na Tabela 11.

6.3.1 Filtragem *Wavelet*

Foi utilizada a modelagem 1 (Tabela 11) na estimativa dos intervalos de confiança e previsão, de onde se determina uma previsão combinada a partir das séries bootstrapadas, sendo esta última previsão comparada com resultados de outros métodos preditivos.

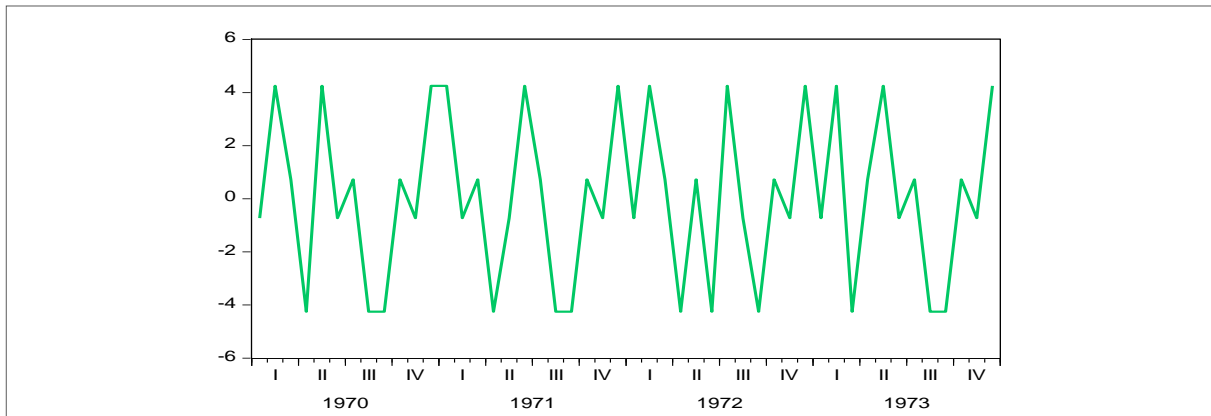
A série de vazões (492 termos) foi dividida em amostra de treino e validação (totalizando 443 termos) e teste com 49 observações. Aos 443 termos (treino e validação) agregaram-se mais 69, resultantes de uma previsão 69 passos a frente por RNA. Desta forma, a série a ser filtrada se encontra com 512 termos. No procedimento do encolhimento *wavelet* (EW) da série de vazões observadas y utilizou-se na decomposição ortogonal a base haar em nível $p = 2$, com limiar universal e regra de limiarização duro (Tabela 11). Com o EW da série $\{y_t\}_{t=1}^{512}$ obteve-se as séries tendência $\{y_{t,EW}\}_{t=1}^{512}$, supostamente menos ruidosa que a série observada e, o resíduo wavelet $\{y_{t,R}\}_{t=1}^{512} = \{y_t\}_{t=1}^{512} - \{y_{t,EW}\}_{t=1}^{512}$. Estas três séries estão representadas na Figura 29. A fim de ilustrar a escala de valores na série y_R , a Figura 30 traz um detalhe (primeiros quatro anos) desta série.

FIGURA 29 - GRÁFICOS DAS SÉRIES DE VAZÕES, TENDÊNCIA E RESÍDUOS



FONTE: O autor (2015).

FIGURA 30 - GRÁFICO DOS RESÍDUOS EW

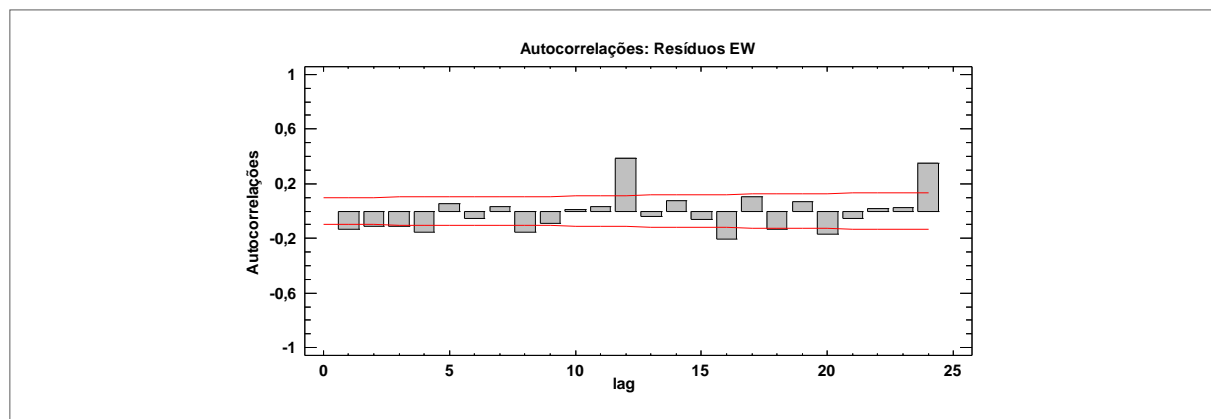


FONTE: O autor (2015).

Testes estatísticos, ao nível de 5% de significância, indicaram que o resíduo (y_R) proveniente do EW da série y não pode ser classificado como ruído branco, de forma que é passível de modelagem.

O gráfico da *função de autocorrelação* (FAC) de y_R representado na Figura 31 mostra alguns picos fora do intervalo onde as autocorrelações são consideradas estatisticamente nulas, além disso, o teste de Box- Pierce apresentou p-valor igual a zero, sendo mais um indicativo da presença de autocorrelação na série de resíduos.

FIGURA 31 - FAC DA SÉRIE DE RESÍDUOS ORIUNDOS DO EW



FONTE: O autor (2015).

Analisando os p-valores registrados na Tabela 12 e nível de 5%, conclui-se que a hipótese nula do teste BDS deve ser rejeitada, indicando a presença de estruturas de auto dependência linear e/ou não linear na série de resíduos.

TABELA 12 - TESTE DE INDEPENDÊNCIA ESTATÍSTICA BDS

Dimensão	Estatística BDS	Estatística z	p-valor
2	-0,0029	-1,33	0,1811
3	-0,0078	-2,28	0,0224
4	-0,0102	-2,58	0,0098
5	-0,0105	-2,59	0,0095
6	-,0,0750	-1,99	0,0462

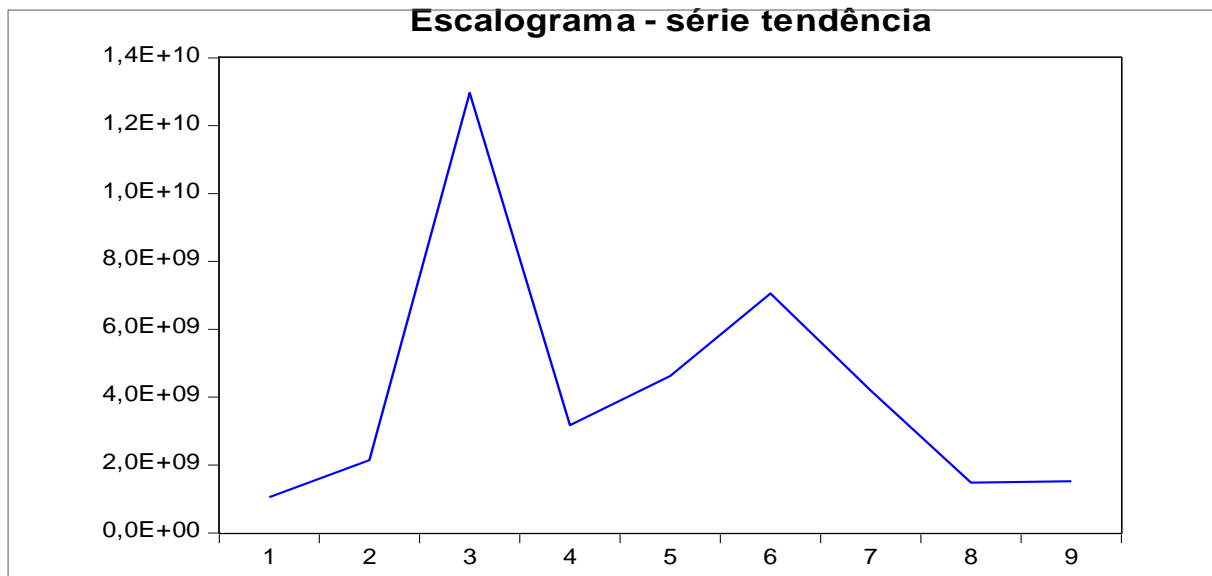
FONTE: O autor (2015).

6.3.2 Decomposição *Wavelet*

A série de tendência e resíduos foram decompostas a partir da base db 40 em nível $p = 9$. Como dissertado na seção 4.2, procedeu-se o cálculo da média da série $\{y_{t,EW}\}_{t=1}^{512}$, chegando-se ao valor $11810 \text{ m}^3/\text{s}$, para em seguida obter a série centrada em zero $\{Y_{t,EW}\}_{t=1}^{512} = \{y_{t,EW}\}_{t=1}^{512} - 11810$. Então, fez-se, a partir da base db 40 em nível $p = 9$, a DW das séries $\{Y_{t,EW}\}_{t=1}^{512}$ e $\{y_{t,R}\}_{t=1}^{512}$.

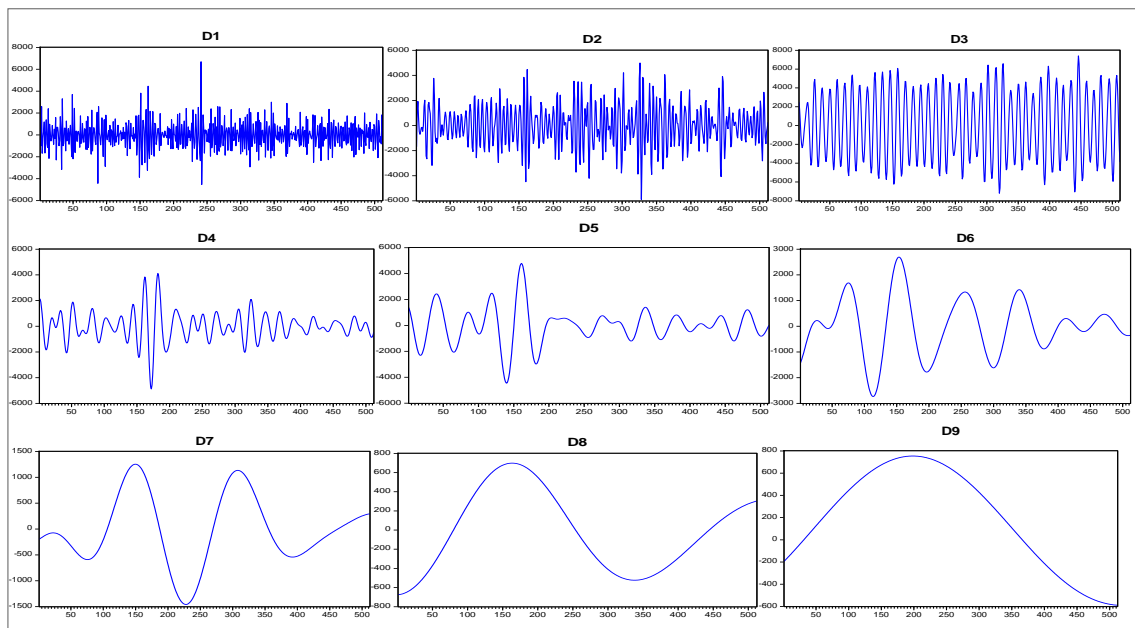
A fim de constituir as séries temporais $X_{t,EW}$ e $Z_{t,EW}$, construiu-se o escalograma (seção 4.2) para os 9 níveis da DW. A análise do escalograma (Figura 32) e uma inspeção visual nos gráficos das componentes wavelet de detalhes (Figura 33), possibilitam a escolha do nível limite π na formação da série $Z_{t,EW}$. O escalograma (Figura 32) apresenta picos de energia nos níveis 3 e 6, sendo representantes das componentes cíclicas de altas e baixas frequências, respectivamente. Observa-se nos gráficos da Figura 33 que as frequências diminuem do nível 1 (D_1) para 9 (D_9). Considerando a análise dessas figuras, optou-se por $\pi = 3$, formando-se as séries $\{Z_{t,EW}\}_{t=1}^{512} = D_1 + D_2 + D_3$ e $\{x_{t,EW}\}_{t=1}^{512} = D_4 + \dots + D_9 + A_9$. Eliminando-se os últimos 69 instantes dessas séries e somando-se a média $11810 \text{ m}^3/\text{s}$ a $\{x_{t,EW}\}_{t=1}^{512}$, chegam-se as séries $\{X_{t,EW}\}_{t=1}^{443}$ e $\{Z_{t,EW}\}_{t=1}^{443}$ utilizadas no treino e validação.

FIGURA 32 - ESCALOGRAMA – COEFICIENTES DE DETALHES NA DW DA SÉRIE TENDÊNCIA.



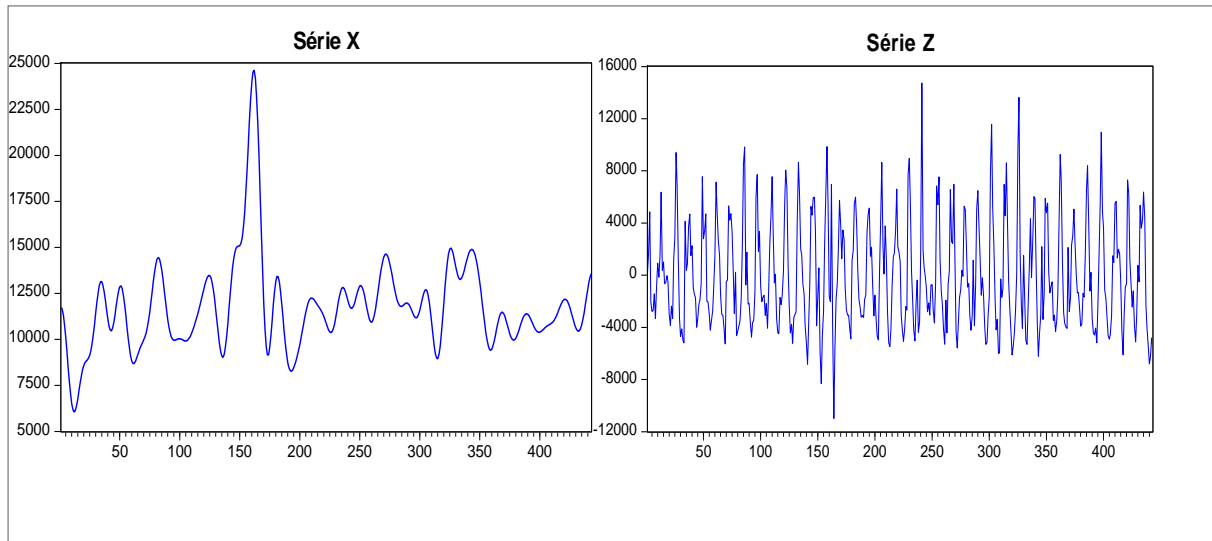
FONTE: O autor (2015).

FIGURA 33 - GRÁFICOS DAS COMPONENTES WAVELET DE DETALHES- SÉRIE TENDÊNCIAS.

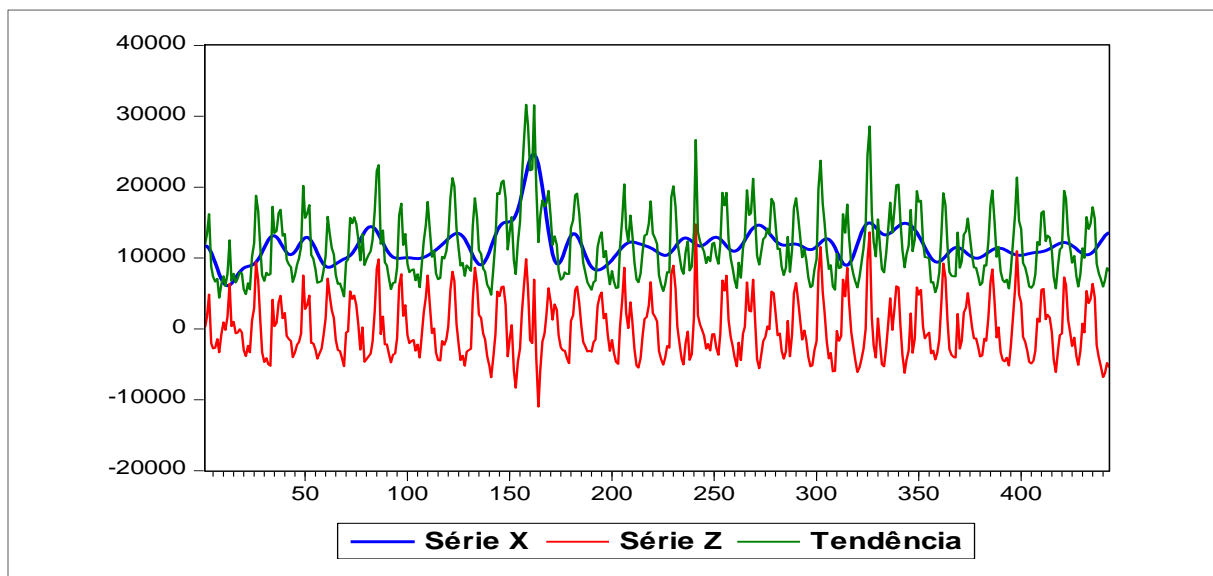


FONTE: O autor (2015).

Na Figura 34 podem ser observados os gráficos das séries temporais $\{X_{t,EW}\}_{t=1}^{443}$ e $\{Z_{t,EW}\}_{t=1}^{443}$. A Figura 35, por sua vez, traz os gráficos das séries X , Z e tendência.

FIGURA 34 - GRÁFICOS DAS SÉRIES TEMPORAIS X E Z – TENDÊNCIA.

FONTE: O autor (2015).

FIGURA 35 - GRÁFICOS DAS SÉRIES TENDÊNCIA, X E Z . $TENDÊNCIA \cong X + Z$.

FONTE: O autor (2015).

Utilizou-se procedimento semelhante para a determinação das séries $\{X_{t,R}\}_{t=1}^{443}$ e $\{Z_{t,R}\}_{t=1}^{443}$, observando que neste caso a série $\{y_{t,R}\}_{t=1}^{T+q}$ é submetida a DW sem que seja subtraída a sua média (veja seção 4.2).

6.3.3 Modelagem por RNA: previsões pontuais

As modelagens por RNA *feedforward* foram realizadas através do *software* Matlab 8.0, num total de quatro redes, sendo duas para as séries X e Z oriundas da

decomposição ortogonal da série tendência e duas para as séries X e Z provenientes da decomposição da série de resíduos EW. As redes eram constituídas por uma camada escondida com função de ativação tangente hiperbólica sigmoidal (*tansig*) ou linear (*purelin*) e uma camada de saída com um neurônio e função de ativação linear, as melhores configurações encontradas estão expostas no Quadro 3.

QUADRO 3 - MAPE E CONFIGURAÇÕES DAS RNA USADAS NA MODELAGEM DAS SÉRIES X E Z PROVENIENTES DAS SÉRIES TENDÊNCIA E RESÍDUOS EW.

Séries oriundas da filtragem EW	Séries oriundas da DW	Função de ativação (camada oculta)	Nº de neurônios na camada escondida	Janela de entrada de dados	MAPE: amostra de validação
Tendência	Série X	<i>purelin</i>	10	14	0,0028%
	Série Z	<i>tansig</i>	10	12	56%
Resíduos EW	Série X	<i>purelin</i>	12	14	0,92%
	Série Z	<i>tansig</i>	12	12	304%

FONTE: O autor (2015).

Na Tabela 13 estão as previsões das séries tendência ($y_{t,EW}$), resíduos EW ($y_{t,R}$) e vazões mensais (y_t) para os meses do último ano (2010). A última coluna (\hat{y}_t) desta tabela resulta da soma das previsões $\hat{y}_{t,EW}$ e $\hat{y}_{t,R}$.

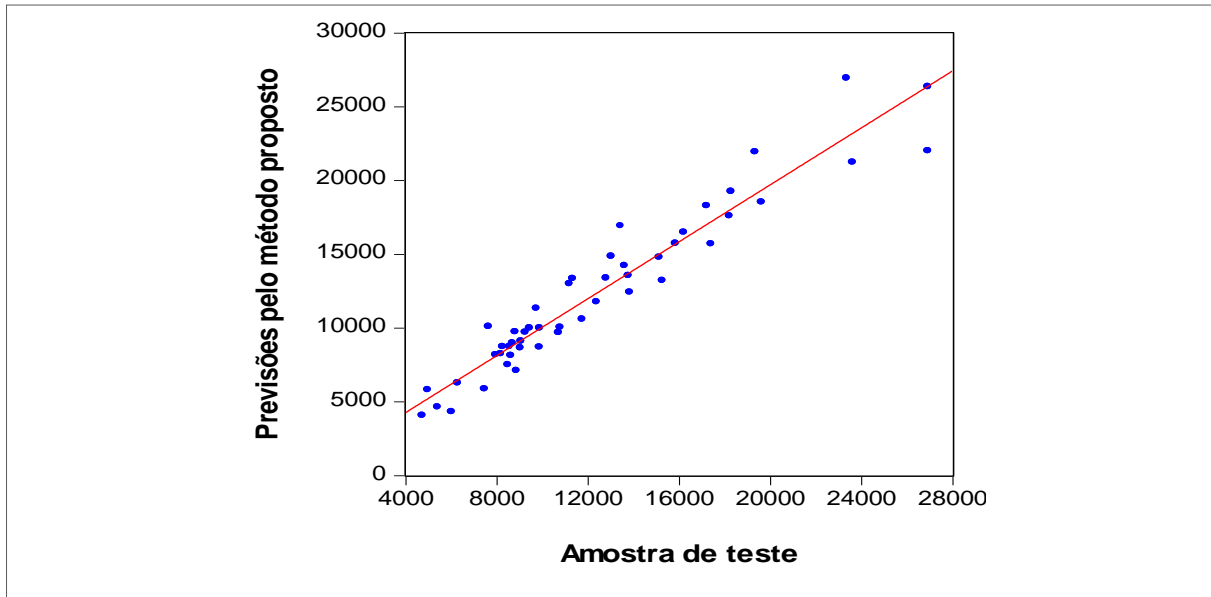
TABELA 13 - PREVISÕES PARA AS 12 ÚLTIMAS OBSERVAÇÕES DA AMOSTRA DE TESTE – VAZÕES MENSAIS POSTO 266-ITAIPU

Meses/2010	y_t	$y_{t,EW}$	$y_{t,R}$	$\hat{y}_{t,EW}$	$\hat{y}_{t,R}$	\hat{y}_t
Jan	23603,9	23599,6	4,25014	22414,2	1,46518	22415,7
Fev	23343,1	23343,9	-0,72921	22782,8	-0,35682	22782,4
Mar	17203,5	17202,8	0,72921	19386,7	1,83277	19387,9
Abr	16189,6	16193,8	-4,25014	15158,4	-1,86045	15156,6
Mai	11319,6	11315,4	4,25014	12210,1	2,53669	12212,6
Jun	9057,6	9058,4	-0,72921	9758,9	-2,02430	9756,2
Jul	78941,8	7941,1	0,72921	7296,2	-0,64539	7295,5
Ago	6268,7	6272,9	-4,25014	6710,3	-3,51668	6706,7
Set	5398,6	5402,9	-4,25014	5319,9	-2,80421	5317,2
Out	9239,9	9239,2	0,72921	6188,2	1,03508	6189,3
Nov	8675,8	8676,5	-0,72921	9832,7	-3,63876	9829,0
Dez	13419,8			11739,9	0,67207	11740,6

FONTE: O autor (2015).

O gráfico de dispersão (Figura 36) da amostra de teste versus previsão pelo método proposto, o coeficiente de correlação (0,97) entre esses dois conjuntos de dados e o MAPE (9,56) na amostra de teste são indicativos da acuracidade das previsões obtidas com o método proposto.

FIGURA 36 - DIAGRAMA DE DISPERSÃO ENTRE A AMOSTRA DE TESTE E PREVISÕES PELO MÉTODO PROPOSTO.



FONTE: O autor (2015).

6.3.4 Intervalo de confiança por *Bootstrap*

O processo *bootstrap* foi utilizado segundo a versão reamostragem residual (seção 2.6), que foi aplicada às séries do Quadro 3 (seção 6.3.3). A quantidade de reamostragens obtidas foi igual a 20 ($B = 20$), número mínimo sugerido por Tibshirani (1996). As séries *bootstrapadas*, 20 para cada série do Quadro 3, foram modeladas segundo o método proposto para previsões pontuais (resumido na Figura 15). Para cada configuração de rede utilizada foram realizadas dez simulações, escolhendo-se a rede com menor erro quadrático médio na amostra de validação, observando que as configurações das RNAs utilizadas na modelagem das séries *bootstrapadas* são semelhantes às mencionadas no Quadro 3, consequência da proximidade das séries desse quadro e suas previsões, levando a valores relativamente pequenos para os resíduos usados na montagem das séries *bootstrapadas*.

Considere o caso da série de tendência $y_{t,EW}$, que após decomposição *wavelet* passa a ser representada pela soma das séries temporais $X_{t,EW}$ e $Z_{t,EW}$. De posse das suas previsões ($\hat{X}_{t,EW}$ e $\hat{Z}_{t,EW}$), obteve-se o conjunto das séries *bootstrapadas* $X_{t,EW}^{*1}, \dots, X_{t,EW}^{*B}$ e $Z_{t,EW}^{*1}, \dots, Z_{t,EW}^{*B}$ e suas previsões $\hat{X}_{t,EW}^{*1}, \dots, \hat{X}_{t,EW}^{*B}$ e $\hat{Z}_{t,EW}^{*1}, \dots, \hat{Z}_{t,EW}^{*1}$. No passo seguinte foram obtidas as médias ponderadas $\bar{\hat{X}}_{t,EW}^*$ e $\bar{\hat{Z}}_{t,EW}^*$.

e com elas as estimativas dos erros padrões $\hat{\sigma}_{X_{t,EW}}$ e $\hat{\sigma}_{Z_{t,EW}}$ (conforme descrito na seção 4.4). Por fim, o intervalo de confiança (IC) oriundo da modelagem da série de tendência, que resulta da soma dos ICs (inequações (56) e (57)) provenientes das modelagens de $X_{t,EW}$ e $Z_{t,EW}$.

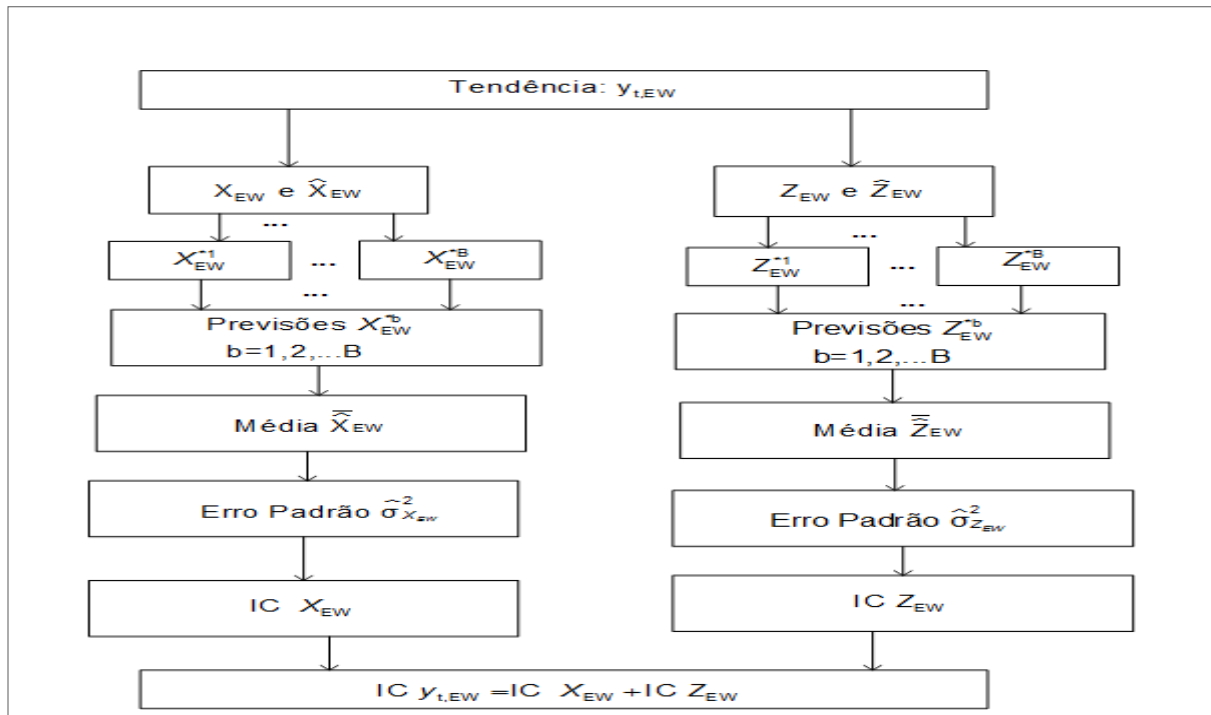
$$\bar{\hat{X}}_{t,EW}^* - t_{1-\frac{\alpha}{2},gl} \hat{\sigma}_{X_{t,EW}} \leq f_X \leq \bar{\hat{X}}_{t,EW}^* + t_{1-\frac{\alpha}{2},gl} \hat{\sigma}_{X_{t,EW}}, \quad (56)$$

$$\bar{\hat{Z}}_{t,EW}^* - t_{1-\frac{\alpha}{2},gl} \hat{\sigma}_{Z_{t,EW}} \leq f_Z \leq \bar{\hat{Z}}_{t,EW}^* + t_{1-\frac{\alpha}{2},gl} \hat{\sigma}_{Z_{t,EW}}, \quad (57)$$

O procedimento descrito no último parágrafo está representado no fluxograma da Figura 37. Para o cálculo do IC referente à modelagem da série de resíduos *wavelet* ($y_{t,R}$) utilizou-se o mesmo procedimento e, a soma deste com o IC de $y_{t,EW}$ fornece o IC da série observada, conforme equação (58). Observando que a previsão de y_t resulta da soma das médias $\bar{\hat{X}}_{t,EW}^*$, $\bar{\hat{Z}}_{t,EW}^*$, $\bar{\hat{X}}_{t,R}^*$ e $\bar{\hat{Z}}_{t,R}^*$.

$$IC\ y_t = IC\ y_{t,EW} + IC\ y_{t,R}. \quad (58)$$

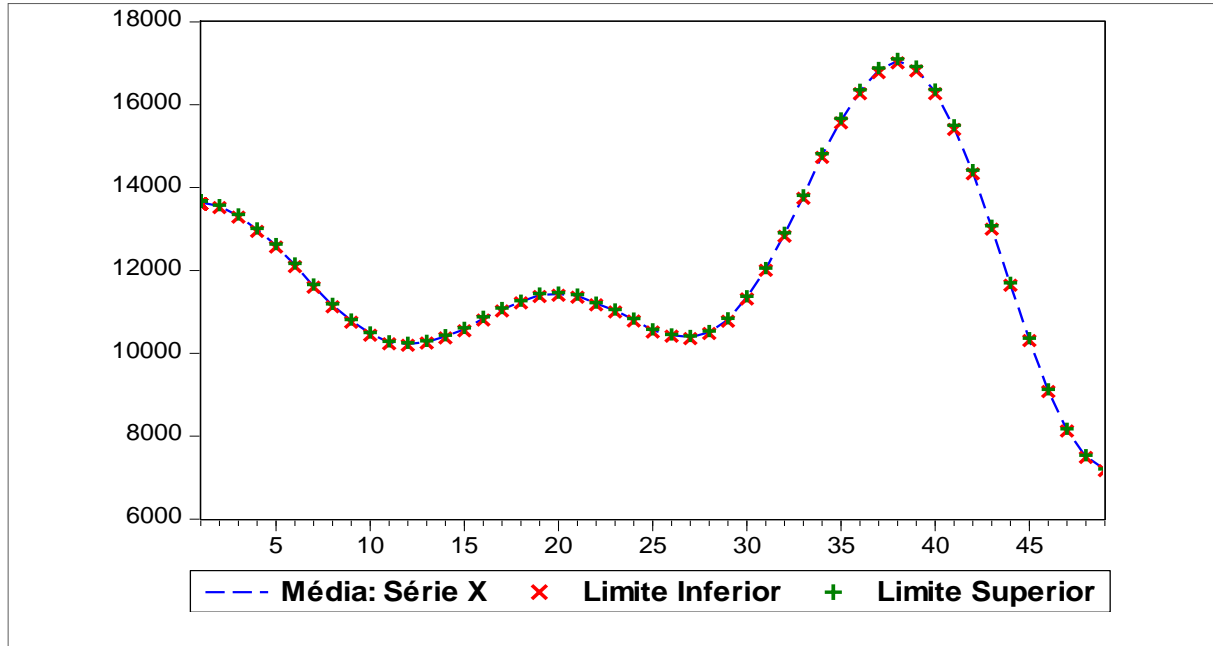
FIGURA 37 - FLUXOGRAMA COM AS ETAPAS DO PROCEDIMENTO *BOOTSTRAP* NA ESTIMATIVA DO IC.



FONTE: O autor (2015).

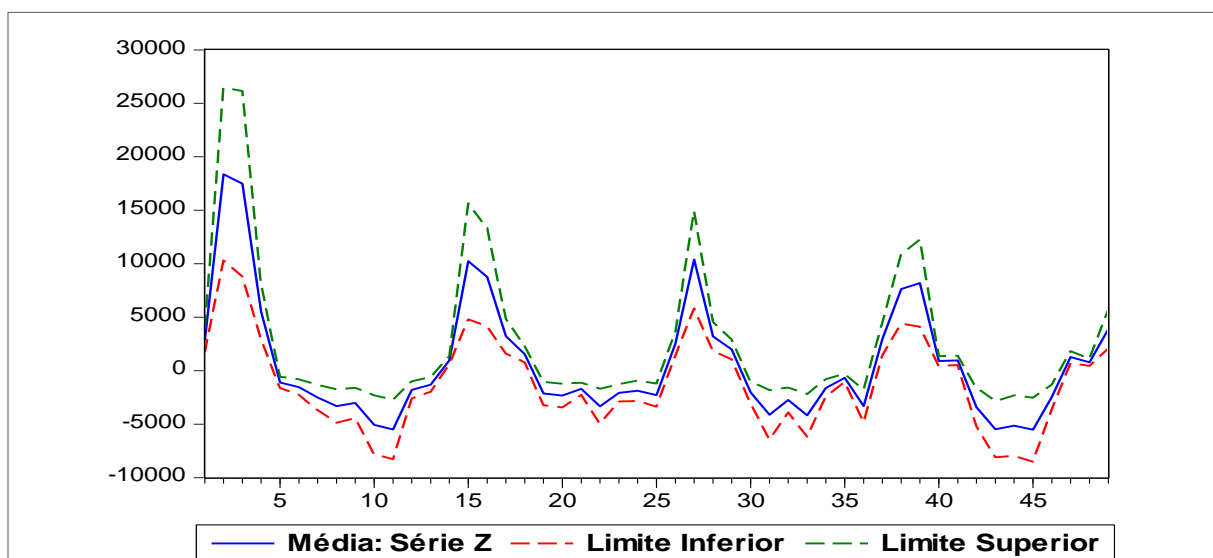
As Figuras 38 e 39 ilustram os ICs obtidos para X e Z relativas à série tendência $y_{t,EW}$. Nelas, as legendas: “Média: Série X” e “Média: Série Z” representam, respectivamente, as previsões de $X_{t,EW}$ e $Z_{t,EW}$ após as reamostragens *bootstrap*. Na figura 40, estão representadas as previsões das vazões (oriundas do procedimento *bootstrap*) e respectivo intervalo de confiança.

FIGURA 38 - IC PREVISÕES DA SÉRIE X: TENDÊNCIA



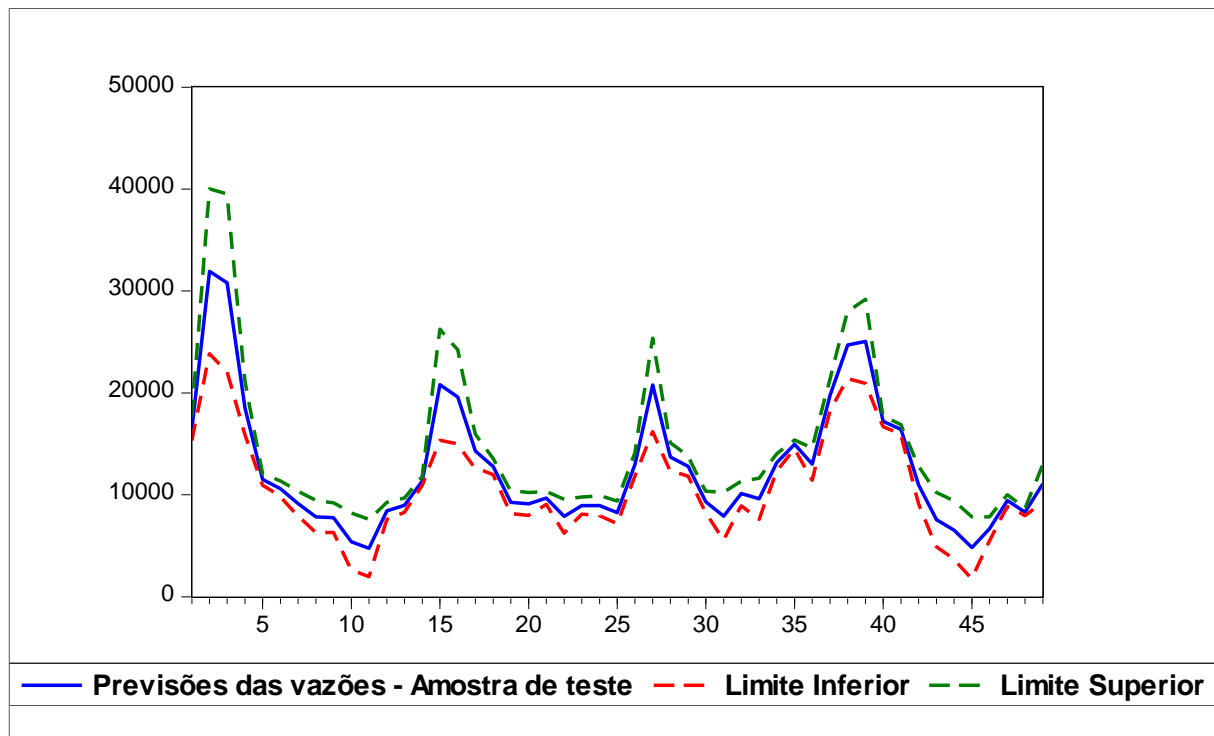
FONTE: O autor (2015).

FIGURA 39 - IC PREVISÕES DA SÉRIE Z: TENDÊNCIA



FONTE: O autor (2015).

FIGURA 40 - IC PREVISÕES DA SÉRIE DE VAZÕES – AMOSTRA DE TESTE



FONTE: O autor (2015).

6.3.5 Previsões Intervalares

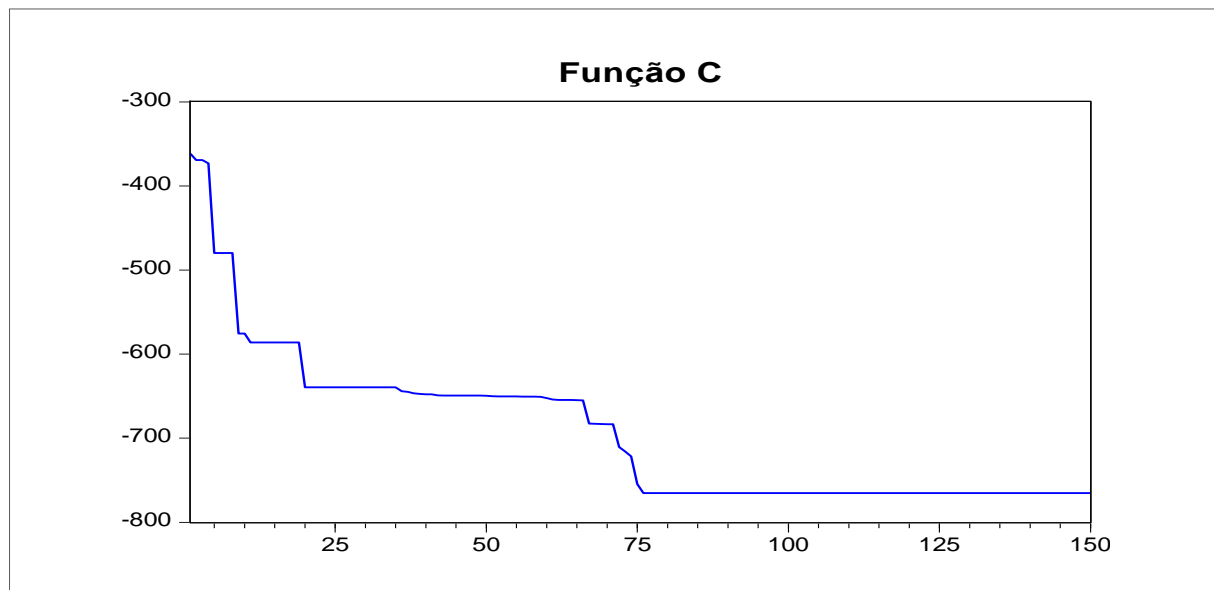
Tendo a estimativa da variância da regressão $\hat{\sigma}_g^2$ (obtida na seção anterior) no modelo $y_t = f(\mathbf{x}, \mathbf{w}) + e_t$, sendo $g(\mathbf{x}, \hat{\mathbf{w}})$ uma aproximação de $f(\mathbf{x}, \mathbf{w})$, o passo seguinte consiste na determinação da variância do ruído e_t a partir da modelagem do resíduo $R^2(\mathbf{x}) = \max\{[y(\mathbf{x}) - \hat{y}_t^*(\mathbf{x})]^2 - \hat{\sigma}_g^2, 0\}$ (seção 4.5). A rede utilizada no ajuste de R^2 era constituída de uma camada oculta e uma de saída, com 10 neurônios na primeira e um na segunda. A função de transferência utilizada na camada oculta foi a tangente hiperbólica e devido à necessidade de se obter resultados não negativos, optou-se pela função exponencial na camada de saída. Entre os tamanhos de janela de entrada testados, verificou-se que os melhores resultados na minimização da função custo C (equação 54, seção 4.5) foram alcançados com padrões de comprimento igual a oito. A otimização da função C foi executada a partir da meta-heurística PSO, cujos parâmetros estão listados na Tabela 14.

TABELA 14 - PARÂMETROS META-HEURÍSTICA PSO

Parâmetros PSO	Valores
Coeficiente cognitivo individual c_1	0,2
Coeficiente cognitivo social c_2	0,3
Taxa de Inércia	0,3
Número de iterações	300
Número de partículas	100

FONTE: O autor (2015).

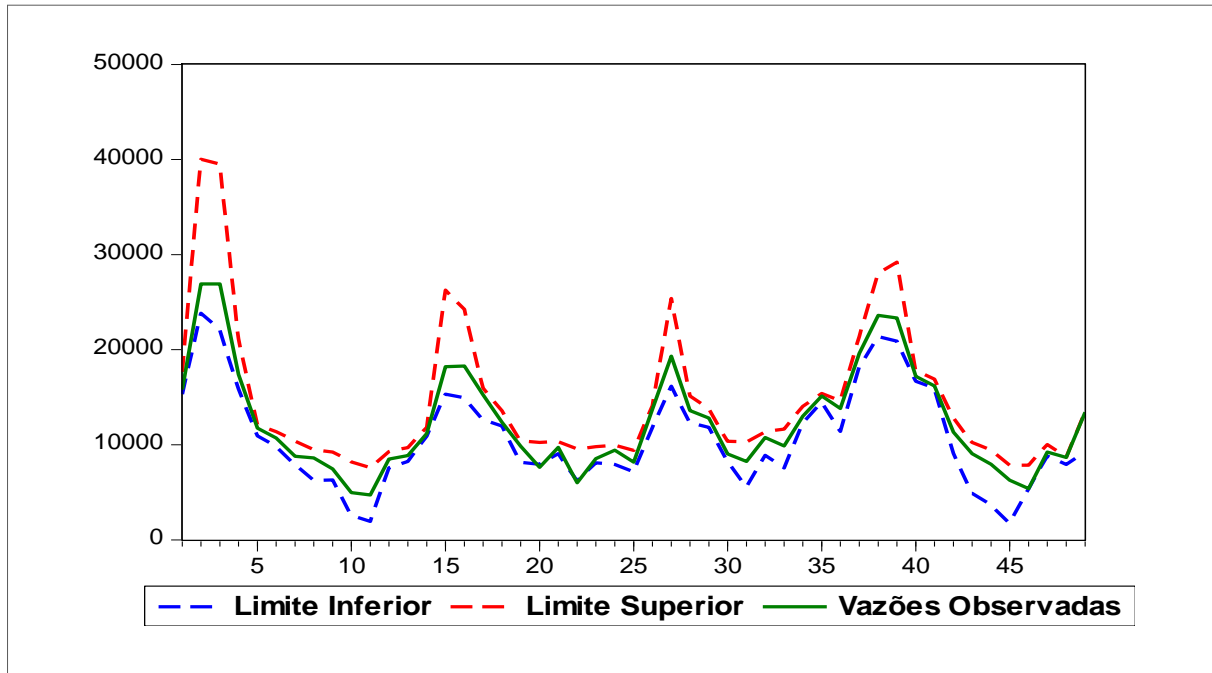
Para o critério de parada no procedimento de otimização foi estabelecido o número de iterações ou a estabilização no valor da função custo C , que na melhor simulação ocorreu a partir da septuagésima quinta iteração, como mostra o gráfico da Figura 41.

FIGURA 41 - COMPORTAMENTO DA FUNÇÃO C DURANTE O PROCESSO DE OTIMIZAÇÃO

FONTE: O autor (2015).

Somando o desvio padrão do ruído, obtido com a otimização de C , ao desvio padrão do modelo (seção anterior), pode-se montar o intervalo de previsão para a série de vazões, como pode ser visto no gráfico da Figura 42.

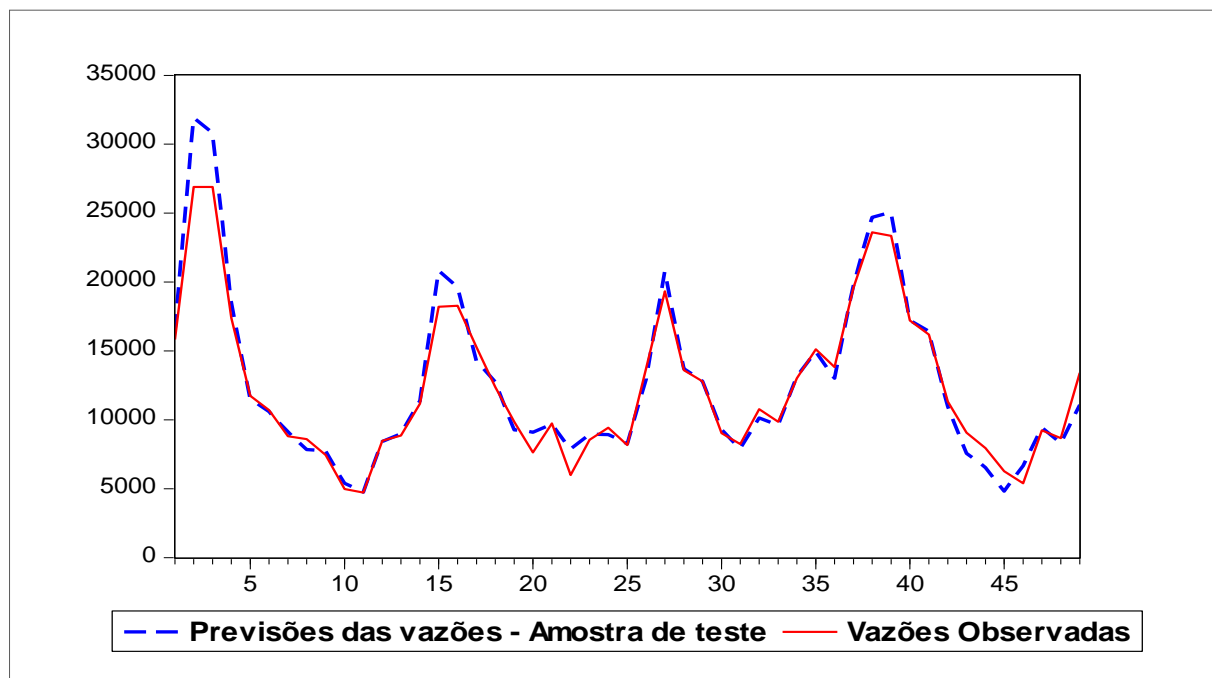
FIGURA 42 - INTERVALO DE PREVISÃO PARA A SÉRIE DE VAZÕES – AMOSTRA DE TESTE



FONTE: O autor (2015).

A Figura 43, por sua vez, traz os gráficos das previsões das vazões na amostra de teste e as respectivas vazões observadas.

FIGURA 43 - GRÁFICOS DAS VAZÕES OBSERVADAS (AMOSTRA DE TESTE) E SUAS PREVISÕES PELO MÉTODO PROPOSTO



FONTE: O autor (2015).

6.4 COMPARATIVO ENTRE OS MÉTODOS

A Tabela 15 informa o MAPE e o MAE correspondentes às previsões feitas um passo à frente da amostra de teste (49 observações) retirada da série de vazões, onde se pode observar uma superioridade do método proposto em relação aos outros métodos (ARIMA, RNA, *Wavelet*-ARIMA e *Wavelet*-RNA).

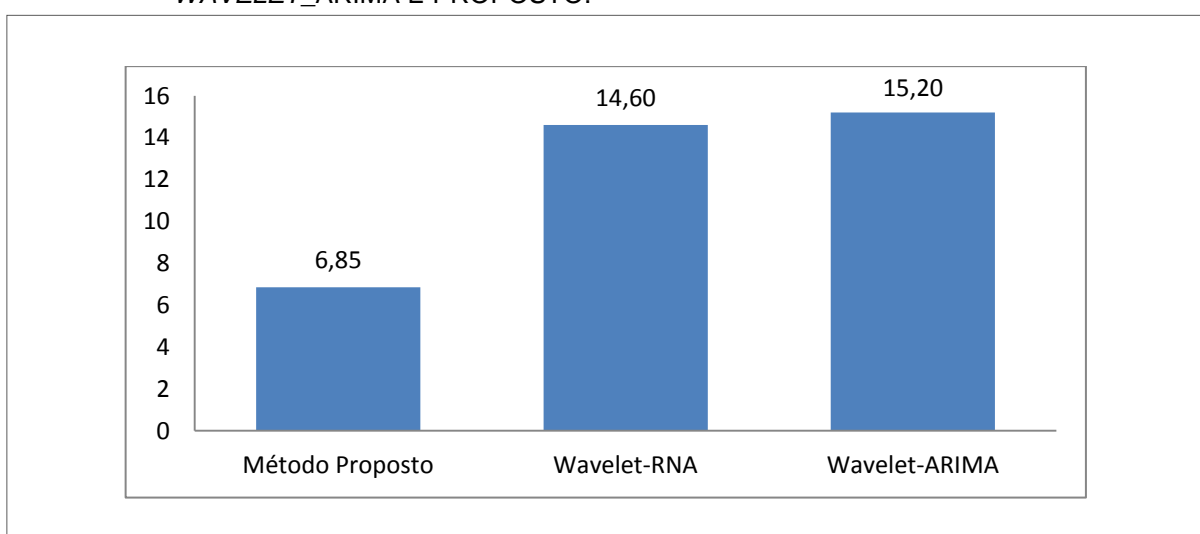
TABELA 15 - MAPE E MAE PARA AMOSTRA DE TESTE RETIRADA DA SÉRIE DE VAZÕES POSTO 266-ITAIPU.

Método	Medidas de acurácia	
	MAPE	MAE
Método proposto	6,85%	829,01
<i>Wavelet</i> _RNA	14,60%	1870,56
<i>Wavelet</i> _ARIMA	15,20%	1932,14
ARIMA	18,01%	2104,39
RNA	18,90%	2112,64

FONTE: O autor (2015).

Os melhores resultados foram obtidos pelos métodos *Wavelet*-RNA, *Wavelet*-ARIMA e proposto, sendo o último o mais eficiente com MAPE 54,93% menor em relação ao *Wavelet*-ARIMA e 53,08% em relação ao *Wavelet*-RNA, como mostra o gráfico comparativo da Figura 44.

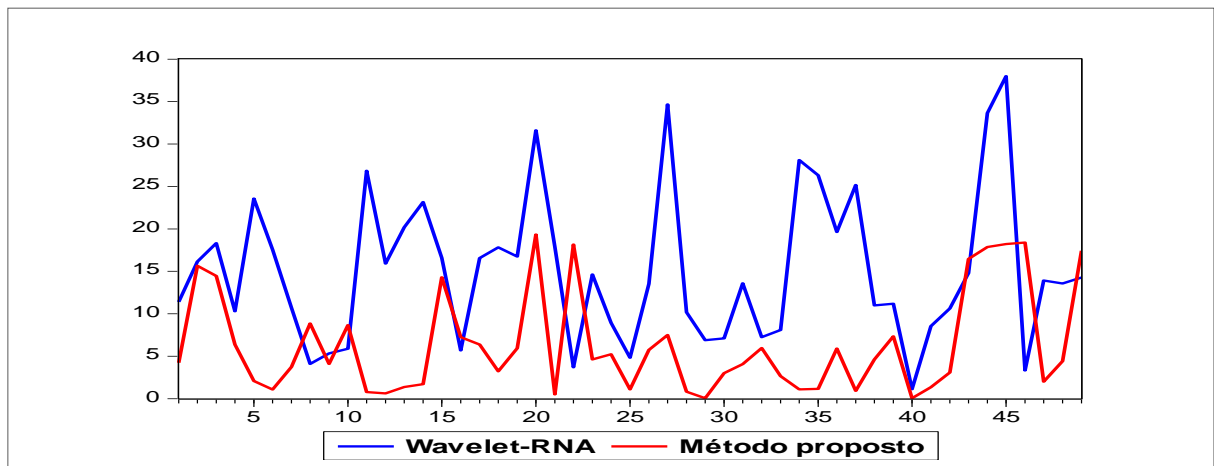
FIGURA 44 - MAPE AMOSTRA DE TESTE PARA OS MÉTODOS *WAVELET*_RNA, *WAVELET*_ARIMA E PROPOSTO.



FONTE: O autor (2015).

Na Figura 45, tem-se a comparação entre o método proposto e *Wavelet*-RNA (segunda abordagem com melhor resultado). Nota-se que, em todos os 49 instantes, na amostra de teste, o método proposto apresenta evolução temporal dos valores dos erros percentuais absolutos (APE) mais estáveis que os dos métodos *Wavelet*-RNA. Nesse método, os valores do APE violaram em vários instantes a faixa dos 20%, enquanto que em nenhum instante isso ocorre no método proposto. Além disso, na maioria dos instantes, o método proposto apresenta valores de APE inferiores ao *Wavelet*-RNA.

FIGURA 45 - APE NA AMOSTRA DE TESTE PARA OS MÉTODOS *WAVELET*_RNA E PROPOSTO.



FONTE: O autor (2015).

7 CONCLUSÕES

Revisando a literatura especializada na análise e previsão de séries temporais, constatou-se a existência de uma crescente tendência no uso de combinações de métodos base de previsões e abordagens de pré-processamento, tais como filtragem e expansões ortogonais. Uma breve inspeção em bancos de dados de trabalhos publicados na área permite acesso a várias propostas nessa linha. Ainda não existe uma combinação de métodos preditivos que seja considerada a melhor na projeção de séries temporais (WALLIS, 2011). No campo da filtragem de séries de tempo, os autores Hassani e Zhigljavsky (2009); Menezes *et al.* (2014); Golyandina *et al* (2001); Donoho e Jonhstone (1994); Donoho e Jonhstone (1995); Elsner e Tsonis (1996) verificaram que métodos de filtragem são imprescindíveis, quando usados adequadamente, na obtenção de ganhos preditivos. Ainda no pré-processamento do sinal, a teoria *wavelet* permite decompor a série em componentes ortogonais que possibilitam um ajuste mais refinado pelos regressores, os autores Karthikeyan e Kumar (2013); Tiwari e Chatterjee (2010); Kisi (2010); Kisi e Cimen (2011); Nalley *et al.*, (2012) e Teixeira Jr (2013) constataram essa característica.

Consequência dessas leituras, o presente trabalho apresentou um método de modelagem de séries temporais que envolve combinação de filtragem e decomposição *wavelet* no pré-processamento dos dados, modelando os resultados deste estágio por meio de redes neurais artificiais.

A pesquisa desenrolou-se por várias etapas, fazendo-se em um primeiro momento previsões sem o uso da filtragem *wavelet*. O seu uso, dentro da proposta apresentada na tese, permitiu ganhos preditivos consideráveis, atestando a sua importância na modelagem de séries temporais, particularmente nas mais ruidosas. Em uma das fases da pesquisa procedeu-se a modelagem de uma série de vazões mensais coletadas no posto 266 (Itaipu). Foram utilizados, além do método proposto, os consagrados métodos ARIMA e RNAs e, as composições *Wavelet*-ARIMA e *Wavelet*-RNA. O erro MAPE nas modelagens por ARIMA e RNA ficou na ordem de 18%. A simples associação deles com a decomposição *wavelet* possibilitou a redução do MAPE para valores em torno de 14%. Mostrando, para

esta série, a eficiência da decomposição *wavelet* no contexto das previsões de séries temporais. Corroboram com esses resultados as conclusões de outros autores já citados anteriormente.

A eficiência do método proposto também foi constatada na modelagem das séries *Canadian Lynx*, *Wolf's Sunpot* e *Exchange rate*. Utilizadas por outros autores com trabalhos publicados em periódicos conceituados. Comparando os resultados obtidos por esses autores aos alcançados pelo método proposto, mostrou ser este último bastante competitivo e eficiente, com altos ganhos preditivos e grande margem nas diferenças entre os erros de previsão. Salientando que, embora o embasamento teórico do método proposto seja matematicamente complexo, a utilização do software Matlab viabiliza a sua utilização, ao passo de poder ser operacionalizado de forma relativamente simples.

Por último, estimou-se as incertezas das previsões realizadas para a série de vazões mensais. São vários os procedimentos encontrados na literatura para este fim, especialmente no caso de regressor neural. Tibshirani (1996) descreveu e comparou em inúmeros exemplos três métodos para estimar o erro padrão de previsões provenientes de redes *multi-layer perceptron*. Utilizou o método Delta, estimador *sandwich* e *bootstrap*, concluiu que este último apresentou o melhor desempenho, capturando de forma mais eficiente a variabilidade do processo em consequência da aleatoriedade dos pesos iniciais das redes *bootstrap*. Tibshirani (1996) utilizou no seu trabalho a média aritmética simples das previsões das séries *bootstrapadas*. Nesta tese foi proposto o uso da média ponderada, procurando com essa estratégia penalizar os resultados com maior erro preditivo. Os resultados obtidos mostraram que fazendo modelagens adequadas das séries *bootstrapadas* podem levar a uma redução dos erros de previsão do modelo inicialmente ajustado no procedimento *bootstrap* residual. Em contrapartida o custo computacional para a execução desse procedimento é bastante alto, onerando o tempo para execução do mesmo. Além do intervalo de confiança, estimado por meio do *bootstrap*, obteve-se o intervalo de previsão. Para tanto, foi necessário a criação de uma rede neural artificial para estimar a variância do ruído. Uma opção bastante usada no treinamento de RNA é o algoritmo *backpropagation* (BP). Mas, é sabido que o algoritmo BP baseia-se no gradiente descendente que depende dos valores iniciais e quase sempre converge para um ponto que não é o ideal. Diferentes algoritmos evolutivos têm sido propostos para melhorar a capacidade da rede neural

(ANGELINE, SAUNDERS e POLLACK, 1994; LIANG, 2007; TANG, QUEK, e NG, 2005). Neste trabalho, a função custo da rede utilizada na estimação da variância do ruído foi minimizada via PSO convencional que apresentou resultados satisfatórios quando comparados com outras meta-heurísticas (Algoritmo Genético (AG) e *Simulated Annealing*). Zhao e Yang (2009) utilizaram uma variação do PSO no treinamento de uma RNA e auferiram melhores resultados que os obtidos por meio dos métodos BP e AG. O Intervalo de previsão (IP) estimado apresentou baixa variação e abrigou todas as vazões observadas na amostra de teste, não sendo os seus limites violados. Mostrando que o procedimento utilizado na estimação do IP, embora com custo computacional elevado, pode apresentar rendimento superior à outras abordagens.

7.1 SUGESTÕES PARA TRABALHOS FUTUROS

Os resultados encontrados neste trabalho reforçam a tendência dos modelos quantitativos de previsão de se fazer uso combinado de técnicas de modelagem e pré-processamento dos dados. Como não há um consenso na literatura a respeito de qual a melhor combinação de métodos preditivos, variações no método proposto merecem ser investigadas. Tal como uma combinação híbrida multiestágios que aproveita a robustez dos modelos Box e Jenkins na modelagem de estruturas lineares com a versatilidade das redes neurais artificiais, combinando linearmente ou não linearmente as previsões oriundas desses regressores.

Nesta tese, o modelo proposto utilizou redes neurais *feedforward*, fazendo-se previsões um passo a frente. Sugere-se adaptar o modelo para previsões vários passos a frente e utilizar outros tipos de redes neurais, sendo as RNAs recorrentes bastante recomendadas. Também pode ser investigada a eficiência do método proposto na modelagem de séries temporais caóticas, como Mackey-Glass e Sistema de Lorenz, por exemplo.

No quesito aplicação na previsão de vazões, a presente pesquisa trabalhou com vazões mensais. As séries de vazões semanais e diárias apresentam comportamento diferenciado em relação às mensais, veja Wang *et al.*, (2006a, 2006b). Seria recomendado pesquisar o comportamento do método na modelagem dessas séries, observando que previsões de curto prazo como as diárias, por

exemplo, são de grande importância no planejamento operacional das usinas hidrelétricas.

Embora os resultados encontrados neste trabalho sejam bastante animadores, não é pretensão do mesmo encerrar as discussões sobre a linha de pesquisa abordada e com certeza pode-se dizer que o tema está aberto a discussões, sendo inúmeras as possibilidades de combinações das técnicas aqui trabalhadas e de outras existentes na literatura e não menos eficientes.

REFERÊNCIAS

ARIÑO, M. A. **Time Series Forecasts via Wavelets: An Application to Car Sales in The Spanish Market**, Discussion Paper 95-30, ISDS, Duke University, 1995.

ABRAMOVICH, F.; BAILEY, T. C.; SAPATINAS, T. **Wavelet analysis and its statistical applications**. Journal of the Royal Statistical Society Series D (The Statistician), v. 49, p.1–29, 2000.

ABREU, T.; ARAÚJO, K.A.M.; LOPES, M.L.M.; LOTUFO, A.D.P. Metodologia híbrida utilizando os modelos ARIMA e redes neurais artificiais para previsão de cargas elétricas. **Anais...** Campina Grande: XIX Congresso Brasileiro de Automática, 2012.

ADHIKARI, R.; AGRAWAL, R. K. A Novel Weighted Ensemble Technique for Time Series Forecasting. **Lecture notes in computer Science**, v.7301, p.38-49, 2012.

ALADAG, C. H.; EGRIOGLU, E.; KADILAR, C. Forecasting Nonlinear Time Series With a Hybrid Methodology. **Applied Mathematics Letters**, v22, p.1467-1470, 2009.

ANGELINE, P. J.; SAUNDERS, G. M.; POLLACK, J. P. An evolutionary algorithm that constructs recurrent neural networks. **IEEE Transactions on Neural Networks**, v.5, p. 54–65, 1994.

AQUINO, R. R. B.; LIRA, M. M. S.; OLIVEIRA, J. B.; CARVALHO J.; M. A.; NETO, O. N.; ALMEIDA, G. J. Application of *Wavelet* and Neural Network Models for Wind Speed and Power Generation Forecasting in a Brazilian Experimental Wind Park. **Anais...** International Joint Conference on Neural Networks. Atlanta, USA, 2009.

BALLINI, R. **Análise e Previsão de Vazões Utilizando Modelos de Séries Temporais, Redes Neurais e Redes Nebulosas**. 2000. 169p. Tese (Doutorado em Engenharia Elétrica e Computação) – UNICAMP, Campinas, 2000.

BARTKIEWICZ, W. Prediction Intervals for Short-Term Load Forecasting Neuro-Fuzzy Models. **Przegląd elektrotechniczny**, p. 284-287, 2012.

BATES, J. M.; GRANGER, C. W. J. The Combining of Forecasts. **Operational Research Quarterly**, v.20, n.4, p. 451-468, 1969.

BEHNAMIAN, J.; GHOMI, S.M.T. F. Development of a PSO–SA hybrid metaheuristic for a new comprehensive regression model to time-series forecasting. **Expert Systems with Applications**, v.37, p. 974–984, 2010.

BOX, G.; JENKINS, G. **Time Series Analysis: Forecasting and Control**. San Francisco: Holden-Day, 1970.

BOX, G.; JENKINS, G.; REINSEL, G. C. *Time Series Analysis, Forecasting and Control*. 3.ed. Englewood Cliffs:Prentice-Hall, 1994. 598p.

BUNN, D. W. Combining forecasts. **European Journal of Operational search**, p. 223-229, 1988.

CARNEY J. G., CUNNINGHAM P., BHAGWAN U., “Confidence and prediction intervals for neural network ensembles”. **Anais...** the International Joint Conference of Neural Networks (IJCNN'99), Washington DC, USA, 1999

CASTRO, L. N. **Computação Natural**. São Paulo: Editora Livraria da Física, 2010.

CAVALIERI, R; RIBEIRO, E. P. Combinação de previsões de volatilidade: um estudo. **Revista Economia**, 2011.

CLEMEN, R. T. Combining forecasts: a review and annotated bibliography. **International Journal of Forecasting**, p. 559-583, 1989.

COLIN, E.C. **Pesquisa operacional**. Rio de Janeiro: editora LTC, 2007.

DAUBECHIES, I. Orthonormal basis of compactly supported *wavelets*. **Comm. Appl. Math.**, p. 909-996, 1988.

DAUBECHIES, I. Ten Lectures on *Wavelets* (C B M S N S F Regional Conference Series in Applied Mathematics), v. 61, **Soc for Industrial & Applied Math**, Philadelphia, 1992.

DEUTSCH, M.; GRANGER, C.W.J.; TERASVIRTA, T. The combination of forecasts using changing weights. **International Journal of Forecasting**, p. 47-57, 1994.

DONONO, D.L.; JOHNSTONE, I.M. Minimax Estimation via *Wavelet* Shrinkage. **The Annals of Statistics**, v.26, n.3, p.879-921. 1998.

DONOHO, D. L.; JONHSTONE, I. M. Ideal Spatial Adaptation by *Wavelet* Shrinkage. **Biometric**, 81,425-455, 1994.

DONOHO, D. L.; JONHSTONE, I. M.; KERKYACHARIAN, G.; and PICARD, D. *Wavelet* Shrinkage: Asymptopia? (with discussion) **Journal Royal Statistical Society. B**, 57:301-369, 1995.

DONOHO, D.L.; JOHNSTONE, I. M. Adapting Unknow Smoothness via *Wavelet* Shrinkage. **Division of Biostatistics/Stanford University Technical Report**, n.161,1995.

ECHER, E.; NORDEMANN,D.J.R.; VIEIRA,L.E.A.;PRESTES,A.P.;FARIA,H.H. O número de manchas solares, índice da atividade do sol. **Revista Brasileira de Ensino de Física**, v.25 n.2, p. 157-163, 2003.

EFRON, B. *Bootstrap* methods: another look at the Jackknife. **The Annals of Statistics**, v.7, n. 1, p. 1-26, 1979.

EFRON, B.; TIBSHIRANI, R.J. *Bootstrap* methods for standart errors, confidence intervals, and other measures of statistical accuracy. **Statistical Science**, v.1, p. 54-77, 1986.

ELSNER J. B.; TSONIS A. A. **Singular Spectrum Analysis: a New Tool in Time Series Analysis**. Plenum Press, New York,1996.

FANG, Y. Forecasting combination and encompassing tests. **International Journal of Forecasting**, p. 87–94, 2003.

FARIA, A. E.; MUBWANDARIKWA, E. (2008). Multimodality on the Geometric Combination of Bayesian Forecasting Models. **International Journal of Statistics and Management System**, p. 1-25, 2008.

FERREIRA, C. C. **Previsão de vazões naturais diárias afluentes ao reservatório UHE Tucuruí utilizada a técnica de redes neurais artificiais**. Goiânia, 117 p. Dissertação (Mestrado Engenharia da Computação) – Universidade Federal de Goiás, Goiânia, 2012.

FIORDALISO, A. A nonlinear forecasts combination method based on Takagi–Sugeno fuzzy systems. **International Journal of Forecasting**, p. 367–379, 1998.

FREEDMAN, D.A.; PETERS, S.C. *Bootstrapping* regression equation: some empirical results. **JASA**, v.79, p. 97-106, 1984.

GOLYANDINA N.; NEKRUTKIN, V.; ZHIGLJAVSKY, A. **Analysis of time series structure: SSA and related techniques**. Chapman & Hall / CRC. London, 2001.

GOOIJER, J. G.; HYNDMAN, R. J. 25 years of time series forecasting. **International Journal of Forecasting**, p.443– 473, 2006.

GRANGER, C. W. J. Invited Review: Combining Forecasts - Twenty Years Later. **Journal of Forecasting**, p. 167-173, 1989.

GRANGER, C.W.J.; RAMANATHAN, R. Improved Methods of Forecasting. **Journal of Forecasting**, v.3, p.197-204, 1984.

GUILHON, L. G. F. **Modelo heurístico de previsão de vazões naturais médias semanais aplicado à usina de Foz do Areia**. Tese de Doutorado em Engenharia. Universidade Federal do Rio de Janeiro, 2002.

GUJARATI, D. N. **Econometria básica**. São Paulo: Pearson Makron Books, 2000.

GUPTA, S.; WILTON, P. C. Combination of Forecasts: An Extension. **Management Science**. v.33, n.3, p.356-372, 1987.

HAAR, A. Zur theorie der orthogonalen funktionen-systeme. **Math. Ann**, p.331-371, 1910.

HAMILTON, J. **Time series analysis**. Princeton University Press, 1994.

HASSANI, H.;ZHIGLJAVSKY, A. Singular Spectrum Analysis: Methodology and Application to Economics Data. **Journal of Systems Science and Complexity**, v. 22, p.372-394, 2009.

HAYKIN, S., **Redes Neurais Princípios e Aplicações**, 2. Ed. Porto Alegre: Bookman, 2001.

HESKES, T. Pratical confidence and predction intervals. **Advances in neural information processing systems**, v.9, p. 466-472, 1997.

HIRPA, F. A. **Hydrologic data assimilation for operational streamflow Forecasting**. Doctoral Dissertations, 153 p.University of Connecticut, 2013.

HOLLAUER, G.; ISSLER, J. V.; NOTINI, H. H. Prevendo o crescimento da produção industrial usando um número limitado de combinações de previsões. **Economia Aplicada**, v. 12, n.2, p. 177-198, 2008.

HWARNG, H. B.; ANG, H. T. A simple neural network for ARMA(p; q) time series. **Singapore: Omega**, v. 29, p. 319-333, 2001.

ISMAIL, S.; SHABRI, A. Time Series Forecasting using Least Square Support Vector Machine for Canadian Lynx Data. **Journal Teknologi**, p.11-15, 2014.

IZMAILOV, A.; SOLODOV, M. **Otimização: condições de otimalidade, elementos de análise convexa e de dualidade**. Rio de Janeiro: IMPA, 2009a.

IZMAILOV, A.; SOLODOV, M. **Otimização: métodos computacionais**. Rio de Janeiro: IMPA, 2009b.

JOHNSTONE, I. M.; SILVERMAN, B. *Wavelet* Threshold Estimators for data with Correlated Noise. **Stanford University Technical Report**, 1996.

KAJITANI, Y.; KEITH, W. H.; McLEOD, A. I. Forecasting Nonlinear Time Series With Feed-Forward Neural Networks: A case Study of Canadian Lynx Data. **Journal of Forecasting**, v.24, p. 105-117, 2005.

KARTHIKEYAN, L.; KUMAR, D. N. Predictability of nonstationary time series using *wavelet* and EMD based ARMA models. **Journal of hydrology**, p. 103-119, 2013.

KHASHEI, M.; BIJARI, M. A new class of hybrid models for time series forecasting. *Expert Systems with Applications*, v. 39, p. 4344-4357, 2012.

KHASHEI, M.; BIJARI, M. A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. **Applied Soft Computing**, v. 11, p. 2664-2675, 2011.

KHASHEI, M.; BIJARI, M. An Artificial Neural Network (p,d,q) Model for Times Forecasting. **Expert Systems With Applications**, v. 37, p. 479-489, 2010.

KENNEDY, J.; EBERHART, R. C. A discrete binary version of the particle swarm algorithm. *Systems, Man, and Cybernetics*, 1997. 'Computational Cybernetics and Simulation', 1997. **Anais... IEEE International Conference on**, v. 5, p. 4104-4108 vol.5, 1997.

KHOSRAVI, A.;NAHAVANDI,S.;SRINIVASAN,D.;KHOSRAVI,R. Constructing Optimal Prediction Intervals by Using Neural Networks and *Bootstrap* Method. **Neural networks and learning systems**, p. 1-6, 2014.

KISI, O. *Wavelet* regression model for short-term streamflow forecasting. **Journal of hidrology**, p. 344-353, 2010.

KISI, O.; CIMEN, M. A *wavelet*-support vector machine conjunction model for monthly streamflow forecasting. **Journal of hidrology**, p. 132-140, 2011.

KUBRUSLY C. S.; LEVAN, N. Abstract *Wavelets* Generated by Hilbert Space Shift Operators. **Adv. Math. Sci. Appl.** 14, 643-660,2006.

LEI, C.; RAN, L. Short-term Wind Speed Forecasting Model for Wind Farm Based on *wavelet* Decomposition DRPT. **Anais...** Nanjing, China, pp 2525-2529, 2008.

LEVAN N.; KUBRUSLY, C. S. A *Wavelet* Time-Shift Detail Decomposition. **Mathematics and Computers in Simulation.** 63, 73-78, 2003.

LEVINE, D. M.;BERENSON,M.L.;STEPHAN,D. **Estatística: Teoria e Aplicações**. 1.ed. Rio de Janeiro: LTC, 2000. 811p.

LIANG, Yi-Hui. Evolutionary neural network modeling for forecasting the field failure data of repairable systems. **Expert Systems with Applications**, v. 33, p.1090–1096, 2007.

LIMA, F. G. **Modelos de previsões de séries temporais financeiras com combinação de filtros de kalman e *wavelets***. Ribeirão Preto, 151p. Tese (Livre Docência em Economia)- Universidade Estadual de São Paulo, Ribeirão Preto, 2011.

LIMA, P. C. *Wavelets*: uma introdução. **Matemática universitária**, n. 33, p. 13-44, 2002.

MACEIRA, M.E.P.;DAMAZIO, J.M.;GHIRARDI,A.O. Periodic ARMA Models Applied to Weekly Streamflow Forecasts. **Anais...** Budapest, IEEE, 1999.

MALLAT, S. **A *Wavelet* Tour of Signal Processing**. Academic Press, San Diego, 1998.

MARTINS, V. L. M. **Comparação de combinação de previsões correlacionadas e não correlacionadas com as suas previsões individuais: um estudo com séries industriais**. Porto Alegre, 100 p. Dissertação de Mestrado em Engenharia de Produção. Universidade Federal do Rio Grande do Sul, Porto Alegre, 2011.

MARTINS, V.L.M.; WERNER, L. Comparação de previsões individuais e suas combinações: um estudo com séries industriais. **Production Journal**, v. 24, n.3, p.618-627, 2014.

MENEZES, L.M.; CASSIANO, K.M.;SOUZA, R.M.; TEIXEIRA JR.,L.A. Modelagem e Previsão de Demanda de Energia com Filtragem SSA. **Revista de Estatística UFOP**, v.III(2), 2014.

MENEZES, L.M.; TEIXEIRA JR, L.A.;CASSIANO,K.M.;SOUZA, R.C.; PESSANHA, J.F. . SSA hierarchical clustering and ARIMA model in forecast of affluent flow. **Anais...**Boston, USA: The 32nd annual international symposium on forecasting, 2012.

MONTGOMERY, D. C.; RUNGER, G. C, **Estatística aplicada e probabilidade para engenheiros**, 2.ed . Rio de Janeiro: LTC, 2003.

MORETTIN, P. A.; TOLOI, L. M. C. **Análise Séries Temporais**. 2ª Ed. ABE Projeto Fisher, Ed. Edgard Blucher, 2006.

MORETTIN, P. A. **Ondas e Ondaletas: Da análise de Fourier à Análise**. São Paulo: EDUSP, 1999.

MORETTIN, P. A. **Wavelets in Statistics**. Institute of Mathematics and Statistics University of São Paulo, Brazil, 2006.

NALLEY, D.; ADAMOWSKI,J.; KHALIL, B. Using discrete *wavelet* transforms to analyze trends in streamflow and precipitation in Quebec and Ontario (1954–2008). **Journal of hydrology**, p. 204-228, 2012.

NEWBOLD, P.; GRANGER, C. W. J. Experience with Forecasting Univariate Time Series and Combination of Forecasts. **Journal Royal Statistical Society**, series A, v.137, n.2, p.131-165, 1974.

NIX D. A.; WEIGEND A. S., Learning local error bars for non – linear regression. **Proceedings of NIPS 7**, pp. 489 – 496, 1995.

PROTÁZIO, J. M. D. **Análise Wavelet Aplicada a Sinais Geofísicos**. Campinas-SP, 150p. Dissertação (Mestrado em Ciências e Engenharia de Petróleo). UNICAMP, Campinas, 2002.

RIVAS, E.; BURGOS, J.C.; AND GARCÍA-PRADA, J.C. **Condition Assessment of Power OLTC by Vibration Analysis Using Wavelet Transform**, IEEE Transactions on Power Delivery, v. 24, p. 687-694, 2009.

SOUZA, R. C.; CAMARGO, M.E. **Análise e Previsão de Séries Temporais: os modelos ARIMA**. Ijuí: SEDIGRAF, 2004.

SOUZA, R. M. **Modelagem de séries periódicas via estruturas PAR(p) utilizando encolhimento wavelet**. Rio de Janeiro, 105 p. Tese (Doutorado Engenharia Elétrica) – Pontifícia Universidade Católica – Rio de Janeiro, 2013.

SOUZA, R.C.; NETO, A.C. A *bootstrap* simulation study in ARMA(p,q) structures. **J. of Forecasting**, v. 15, p. 343-353, 1996.

STONE, L.; HE, D. Chaotic oscillations and cycles in multi-trophic ecological systems. **Journal of theoretical biology**, v. 248, n. 2, p. 382–90, 2007.

SUBBA, R. T.; SABR, M. M. An Introduction to Bispectral Analysis and Bilinear Time Series Models. **Lecture Notes in Statistics**, v. 24, 1984.

SUJJAVIRIYASUP, T. Hybrid ARIMA-Support Vector Machine Model for Agricultural Production Planning. **Applied Mathematical Sciences**, v. 7, n. 57, p. 2833 – 2840, 2013.

TANG, A. M.; QUEK, C.; NG, G. S. GA-TSKfnn: Parameters tuning of fuzzy neural network using genetic algorithms. **Expert Systems with Applications**, v.29, p.769–781, 2005.

TANG, Y.; GHOSAL, S. A consistent nonparametric Bayesian procedure for estimating autoregressive conditional densities. **Computational Statistics & Data Analysis**, v. 51, n. 9, p. 4424–4437, 2007.

TEIXEIRA Jr., L. A. **Combinação SSA-Wavelet de métodos preditivos com ajuste numérico MINIMAX, na geração de previsões e de cenários**. Rio de Janeiro, 114 p. Tese (Doutorado em Engenharia Elétrica). Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2013.

TEIXEIRA Jr., L. A.; PESSANHA, J.F.M.; MENEZAS, M.L.; CASSINAO, K.M.; SOUZA, R.C. Redes neurais artificiais *wavelet* na previsão da radiação solar direta. **Anais...** Rio de Janeiro: Simpósio Brasileiro de Pesquisa Operacional 2012a.

TEIXEIRA Jr., L.A.; MENEZES, M.L.; CASSIANO, K.M.; PESSANHA, J.F.M.; SOUZA, R.C. Modelagem Box e Jenkins, decomposição *wavelet* e programação matemática na geração de previsões. **Anais...** Rio de Janeiro, Brasil: XV Simpósio de Pesquisa Operacional e Logística da Marinha, 2012b.

TEIXEIRA, L.L.; SIQUEIRA, P. H.; TEIXEIRA Jr, L. A. Ajuste de modelos ARMA com as meta-heurísticas PSO e Algoritmo Genético. **Anais...** Bauru, Brasil: XXI Simpósio de Engenharia de Produção, 2014.

TEIXEIRA, L.L.; SIQUEIRA, P. H.; TEIXEIRA Jr, L. A. Previsões de vazões mensais via combinação híbrida ARIMA_NEURAL com encolhimento e decomposição *wavelet*. **Revista Engenharia e Tecnologia**, v. 7, n. 1, p.144-161, 2015.

TERUI, N.; DIJK, H. K. Combined forecasts from linear and nonlinear time series models. **International Journal of Forecasting**, p. 421–438, 2002.

TIBSHIRANI, R. A comparison of some error estimates for neural network models. **Neural Computation**, v.8, p. 152-163, 1996.

TIWARI, M. K.; CHATTERJEE, C. Development of an accurate and reliable hourly flood forecasting model using *wavelet–bootstrap–ANN* (WBANN) hybrid approach. **Journal of Hydrology**, p. 458-470, 2010.

TORRENCE, C.; COMPO, G.P. **A practical guide to wavelet analysis**, **Bulletin of the American Meteorological Society**, vol. 79, no. 1, 61 – 78, 1998.

VALENZUELA, O. Hybridization of intelligent techniques and ARIMA models for time series prediction. **Fuzzy Sets and Systems**, v.159, p. 821-845, 2008.

WALLIS, K. F. Combining forecasts - forty years later. **Applied Financial Economics**, p. 33-41, 2011.

WANG, A.B.W.; VAN GELDER, P.H.A.J.M.; URIJLING, J.K.; MA, J. Forecasting daily streamflow using hybrid ANN models. Amsterdam: **Journal of Hydrology**, v. 324, p. 383-399, 2006a.

WANG, A.B.W.; VAN GELDER, P.H.A.J.M.; URIJLING, J.K.; MA, J.. Testing for nonlinearity of streamflow processes at different time scales. Amsterdam: **Journal of Hydrology**, v. 322, p. 247-268, 2006b.

WERNER, L.; RIBEIRO, J. L. D. Modelo composto para prever demanda através da integração de previsões. **Produção**, v. 16, n. 3, 2006.

WINKLER, R. L.; MAKRIDAKIS, S. The combination of forecasting. **Journal of the Royal Statistical Society**, series A, v.146, p.150-157, 1982.

WONG, C. S.; LI, W. K. On a mixture autoregressive model. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v. 62, n. 1, p. 91–115, 2000.

YAZIZ, S. R. The performance of hybrid ARIMA GARCH modeling in forecasting gold price. **Anais...** Adelaide, Austrália: 20th International Congress on Modelling and Simulation, 2013.

YOON, H.; JUN, S.C.; HYUN, Y.; BAE, G.O.; LEE, K.K. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. **Journal Hydrology**, v.396, p. 128-138, 2011.

YU, J.; WANG, S.; Xi, L. Evolving artificial neural networks using an improved PSO and DPSO. **Neurocomputing**, v.71, p.1054–1060, 2008.

ZHANG, G. P. Time series forecasting using a hybrid ARIMA and neural network model. Atlanta: **Neurocomputing**, v.50, p. 159-175, 2003.

ZHAO, L.; YANG, Y. PSO-based single multiplicative neuron model for time series prediction. **Expert Systems with Applications**, v.36, p.2805–2812, 2009.

ZOU, H.; YANG, Y. Combining time series models for forecasting. **International Journal of Forecasting**, p. 69– 84, 2004.